Major Applied Research Paper No. 1

PROVIDER INCENTIVES AND PRODUCTIVE EFFICIENCY IN GOVERNMENT HEALTH SERVICES

Phase 1:
Review of Concepts, Literature,
and Preliminary Field Work Design

Ricardo Bitran, Ph.D.
Steven Block
Abt Associates Inc.

September 1992

HEALTH FINANCING AND SUSTAINABILITY (HFS) PROJECT

ABT ASSOCIATES INC., PRIME CONTRACTOR
4800 Montgomery Lane, Suite 600
Bethesda, Maryland 20814
Tel: (301) 913-0500 Fax: (301) 652-3916
Telex: 312636

Management Sciences for Health, Subcontractor
The Urban Institute, Subcontractor

AID Contract No. DPE-5974-Z-00-9026-00

ABSTRACT

As Phase I of a three-phase Health Financing and Sustainability project, this paper reviews the relevant concepts and literature and presents the preliminary field work design for research regarding provider incentives and productive efficiency in government health services. In addition to providing a basic conceptual framework of the issues, the authors present empirical and anecdotal evidence from developed and developing countries to assess the potential for provider incentives to improve the efficiency (as well as equity and quality) of health care in developing countries. The discussion includes both monetary and non-monetary, physician and hospital incentives. Finally, the authors consider both the methodological problems and prospective field work activities necessary for completion of phases II (field work) and III (analysis) of the applied research project.

ACKNOWLEDGEMENTS

This paper has benefitted from comments by HFS staff at a seminar held in November of 1991, from written comments by Marty Makinen, Stephanie Wilson, David Parker, David Spencer, and Charles Griffin, and from additional suggestions by the members of HFS's Technical Advisory Group during the meetings of August 1992. We thank Gerard La Forgia for his proposal, included in Chapter 6, to conduct field research in the Dominican Republic.

TABLE OF CONTENTS

LIST OF EXHIBITS

EXECUTIVE SUMMARY

This paper constitutes the first phase of a three-phase HFS major applied research project in the area of "Productive Efficiency: Public Sector Reform" (HFS Applied Research Agenda, 1991). The document provides the foundation for phases two (field work) and three (analysis). A companion HFS applied research Phase I paper entitled "Technical and Economic Efficiency in the Production of Health Services" (Bitran, 1992) complements the definitions and analysis of this document.

This paper presents a conceptual framework for studying the problem of provider behavior, work incentives, and production efficiency in developing country government health systems. The framework is used to support the hypothesis that provider incentives can be adopted to improve productive efficiency in government health services. The model is also used to illustrate how incentives developed to promote greater efficiency can affect the quality of care and the equity of the system.

A review of the literature on provider incentives and health system efficiency is conducted using sources from both developed and developing countries. The review from the industrialized world finds that the literature focuses mainly on economic incentives to physicians and hospitals, the latter primarily through various reimbursement mechanisms. The relevance of the industrialized country literature to the developing country context is established. A small number of documented sources on the experience of developing countries with provider incentives is found. However, interviews with several experts reveal great interest in this line of research.

The final chapter is a preliminary discussion of goals, objectives, and methods of phases two and three of this research. Possible field work opportunities are identified in Niger, the Dominican Republic, and Egypt, and include:

- A study of personnel incentives and performance in government ambulatory health facilities in Niger;
- A study of incentive systems for rural health promoters in the Dominican Republic and the system's effect on recruitment, coverage, and quality; and
- A comparative study of public and private provider incentive systems in hospitals in Egypt.

The study of provider incentives on health worker efficiency, equity, and quality requires the ability both to evaluate these performance measures and to unambiguously establish the effect of incentives on provider performance. Measuring efficiency is a difficult problem due to quality and case mix heterogeneity among providers, as well as to distortions in the prices of production inputs. Efficiency assessment thus calls for methods to measure and control quality, case mix, and price variations among providers. Several techniques are available for this, although they are imperfect. The companion HFS Phase 1 paper on technical and economic efficiency discusses these techniques

and their problems. Isolating the effects of incentives on performance is a more difficult and under-researched problem which will have to be addressed in detail at the final design stage of the field activities.

1.0 INTRODUCTION

Government acts as a major health service provider in most developing countries. Health ministries generally face the challenge of providing nationwide health care at the primary, secondary, and tertiary levels at little or no direct charge to the consumers. In many countries, the capacity of these public health systems falls far short of demand. Indeed, the pressure on government health systems continually increases, as real per capita budgets shrink.

Productive efficiency thus becomes a major concern, as these systems strive to do more with less. Greater efficiency in the production of government health care services could imply either or both a greater quantity or a higher quality of services produced for a given level of inputs. Mechanisms for generating additional resources, especially cost recovery in the form of user fees, are also receiving increasing attention in the policy arena.

Unfortunately, there is growing evidence of inefficiency in government health services in developing countries. In Niger, for example, budget cuts over the past five years have resulted in increasing shortages of pharmaceutical products and medical supplies at Ministry of Health (MOH) facilities.[1] The MOH, however, has kept the number of health personnel constant. This is an inefficient situation, as medical staff who lack drugs and supplies sit idle or provide poor quality care. In Ogun State, Nigeria, a study of efficiency at state-owned and private health facilities found that state facilities use far more non-health workers than is economically efficient.[2] Similarly, a study of productive efficiency in a government-operated hospital in the Dominican Republic found that only 12 percent of all physician-contracted time could be accounted for in patient care activities during the study period.[3]

The challenge, then, is to identify strategies for improving the productive efficiency of government health services. Towards this end, various approaches have been tried. A common approach is the adoption of better management information systems. The notion in this case is that more accurate and timely information regarding the allocation and status of health sector resources can facilitate improved efficiency in their deployment. These strategies have failed to produce expected results when the motivation to use the information produced by the systems is absent.[4] Another family of strategies—the one on which this paper concentrates—involves changing the incentive structure within which health sector providers operate.

---

[1] Information obtained from personal communications with the staff of the USAID-funded Niger Health Sector Support Grant.

[2] Wouters (1990).

[3] Lewis, et al. (1990).

[4] In Niger, for example, new accounting systems have been developed for the Ministry of Health national hospitals, and training and documentation have been provided. The systems have not been used, however, apparently because the personnel fail to see an advantage to their use.

This paper's primary hypothesis is that provider incentives[5] can be adopted to improve the efficiency of government health services in developing countries. A secondary hypothesis is that incentives intended to improve efficiency can also have important implications for equity and quality of care. We provide a conceptual discussion to explore the possible effects of various personnel incentives on efficiency as well as on equity and quality of care.

These incentives may include a broad range of monetary and non-monetary factors. Monetary incentives typically increase the remuneration of health care personnel, though a variety of specific mechanisms exists for establishing the source and level of benefits. Such schemes are often associated with cost recovery programs. For example, health center staffs may be allowed to divide some percentage of health center revenue among them in the form of salary bonuses. Incentives may also be non-monetary in nature. For instance, physicians might be rewarded with specialty medical training for serving a fixed period of time in a remote region. Variations on these themes, as well as an analysis of the implications of these types of incentives, are the main subject of this paper.

The paper is organized as follows:

- Chapter 2 provides a foundation for the subsequent discussion by defining efficiency, equity, and quality in health care.
- Chapter 3 presents a conceptual discussion of how various provider incentives can affect efficiency, and how such incentives may also have secondary effects on quality of care and equity.
- Chapter 4 adds to the foundation for our field research design by reviewing relevant literature and experience with provider incentives in developed countries.
- Chapter 5 presents a similar review for developing countries.
- Chapter 6 is a preliminary discussion of goals, objectives, methods, and costs of phases two and three of this research.

The paper's intended audience is researchers of developing country health services with an interest in health care financing, including public health specialists, economists, and management specialists, and health professionals involved in health services planning and management. While knowledge of basic economic principles will facilitate understanding of this document, it is not necessary.

[5] The terms health worker, provider, and employee are used interchangeably in this document. Provider incentives are incentives given to individual health workers in public or private facilities.

2.0 DEFINITIONS: EFFICIENCY, EQUITY, AND QUALITY IN HEALTH CARE

Before proceeding to an analysis of efficiency and provider incentives, it is necessary to clarify the underlying concepts around which this discussion revolves: efficiency, equity, and quality in the provision of health care.

2.1 TECHNICAL AND ECONOMIC EFFICIENCY

Medical professionals and economists tend to think of efficiency differently; medical professionals are often trained to think in terms of *technical efficiency*, while economists are trained to consider *economic efficiency*.[6] A procedure is technically efficient if production inputs (e.g., labor, drugs, equipment) are combined in a way that yields the maximum feasible output (e.g., outpatient visits, hospitalizations).[7]

In microeconomic terms, a technically efficient production process is one that is placed along the *production possibilities frontier*. This is exemplified in Exhibit 2-1 for a simple medical production process that uses only two production inputs, X_a and X_b (these inputs can be viewed as being doctor and nurse time, or doctor time and drugs, for example).[8] Any point along the production possibilities frontier QQ represents a technically efficient way of combining various quantities of production inputs X_a and X_b to produce the same amount of output, or Q visits. For example, while points 1 and 2 differ in the combination of X_a and X_b (production at 1 is more intensive in X_b than at 2), both permit production of the same quantity Q. Points 1 and 2, like all other points on the frontier QQ, are technically efficient because it is not possible to produce Q with smaller quantities of either X_a or X_b, as those depicted by the line (there is no room for further gain in technical efficiency). Point 3, like all points to the left of the production possibilities frontier, is infeasible; any reduction in the amounts of X_a and X_b from the amounts represented by the frontier necessarily translates into a drop in Q. In contrast, point 4, like all points to the right of the production possibilities frontier, constitutes a technically inefficient way of producing Q visits; technical efficiency can be improved by moving production from 4 to 2, thereby reducing the amount of X_a from X_{a4} to X_{a2}. In sum, one procedure is considered

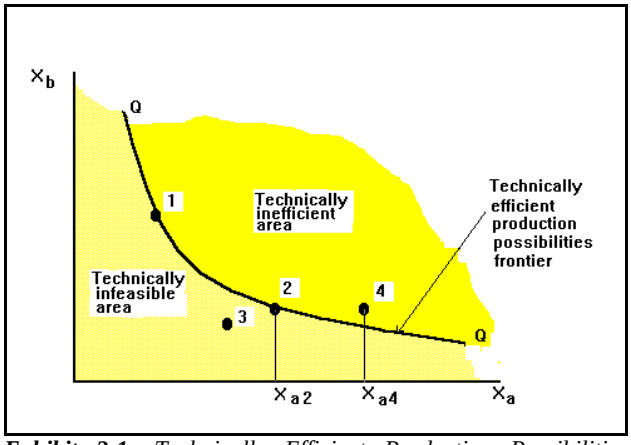

***Exhibit 2-1**: Technically Efficient Production Possibilities Frontier*

[6] Health service researchers in the U.S. use the terms "efficacy" to refer to technical efficiency and "appropriateness" to denote economic efficiency.

[7] Pauly (1970), p.114.

[8] While most health care production processes employ more than two inputs, the visual illustrations used in this document are two-dimensional for graphical convenience. The definitions and analyses presented here are directly generalizable to multiple input production processes.

more technically efficient than another either if it produces the same quantity of output using fewer inputs, or if it produces a greater quantity of outputs using the same resources.

Economic efficiency extends this concept to take into account the relative prices of production inputs. A procedure is economically efficient if inputs are combined to produce a given level of output *at minimum cost*. In general, while many technically efficient alternatives might present themselves to produce a given quantity Q, there is only *one* economically efficient way of doing so.[9]

Exhibit 2-2 helps to illustrate the fundamental difference between technical and economic efficiency. Suppose that the unit prices of X_a and X_b are W_a and W_b, respectively. If the health facility is allocated a budget B_1, then line B_1 represents the facility's budget constraint. The constraint is given by the equation: $B_1 = X_a \cdot W_a + X_b \cdot W_b$. Any point along the budget constraint line, such as points 1 and 3, consumes the total budget B_1. However, point 1 is preferable to 3 because at 1, quantity Q is produced, whereas at point 3, the smaller quantity Q" is produced. Further, of all the technically efficient points along the frontier QQ, point 1 is the most economically efficient way of producing quantity Q. Point 2 is technically as efficient as 1, but is less economically efficient since production at 2 requires a budget of B_2, higher than B_1. Graphically, the economically efficient point (point 1) corresponds to the tangency between the budget constraint and the production possibilities frontier.

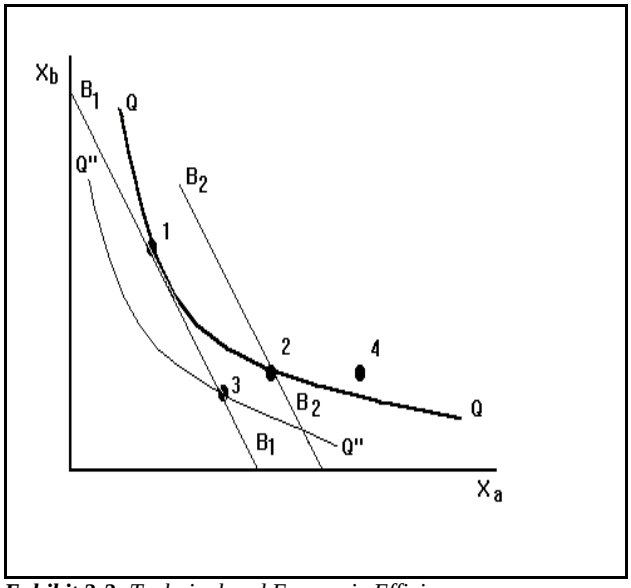

***Exhibit 2-2**: Technical and Economic Efficiency*

In these definitions of technical and economic efficiency, "output" denotes the number of units of medical services or procedures performed. Examples of output measures are: number of hospital discharges, number of surgical interventions, number of children immunized, and number of curative outpatient consultations. In principle, productive efficiency could also be expressed as a function of health "outcomes," instead of outputs. Health outcomes are measures of the population's health status. The improvement of health outcomes is the ultimate goal of health interventions. The number of healthy days of life saved, deaths averted, and illnesses averted are all common examples of health outcome measures as they relate to health interventions.

Establishing a relationship between health output levels and health outcome levels is a complex task, however, because health status is influenced by a large number of variables, such as nutrition, hygiene, medical care, and lifestyle;

[9] There are some unusual production processes which will display more than one economically efficient configuration.

isolating the effect of each variable on health is, therefore, not a straightforward exercise. Because of the difficulty of linking health outputs (e.g., number of condoms sold) with outcomes (e.g., number of cases of HIV/AIDS averted), this paper adopts the more conventional definitions of technical and economic efficiency by using health output as the denominator. Further, to the extent that the quality of the output is constant or can be controlled across providers (see Section 2.3), it can be assumed that, other things being equal, the same number of constant-quality units of output of any pair of providers will produce the same number of health outcomes.

2.2 EQUITY

Equity is a concept that is widely used, though rarely with a common definition. We define equity in public health care to refer to a given person's probability of receiving care of a given quality. We consider a health care system to be equitable if everyone has equal physical and financial access to health care of the same quality for a given set of health problems.[10] If, for example, health facilities are distributed such that urban populations have easier physical access to care, while rural populations may have to travel long distances to receive care, those rural people most likely will have a lower likelihood of receiving care and the system will be inequitable. Differences in the quality of care available to those different populations would also be inequitable. Similarly, if health care is expensive and no provision is made for indigent care, the indigent will most likely have lower economic accessibility to care when perceiving a need for it, thus making the system inequitable.

2.3 QUALITY

Quality of care is of critical concern in the analysis of health care systems. This concern is due to the fact that health care services can vary widely in quality, whereas goods and services in other markets, particularly in commodity markets, tend to exhibit greater quality heterogeneity. Also, quality of care is not easily discerned by consumers of health services because such services are not always traded in competitive markets. We adopt the definition of technical quality used by Wyszewianski et al., who consider quality for care A to be higher than that for care B if care A is likely to make a greater net contribution to the patient's health and well-being than care B. These authors further explain that the expected effect on health is determined in part by whether one type of care is more appropriate than another, and in part by whether the procedure is correctly performed.[11] According to this definition, quality of care has two components: the service's ability to improve the patient's health status and the extent to which providers comply with technical norms of care. We call this definition *technical quality* of care because it alludes to the health effects and technical processes of medical care.

A corollary of the above definition is that if care A and care B have the same quality, but care A is cheaper to produce, then care A is also more

[10] For a useful discussion about the measurement of equity in health, see Musgrove (1986).

[11] Wyszewianski, Thomas, and Friedman (1987), p. 18.

economically efficient. Equivalently, if health care interventions A and B cost the same but care A is of higher quality, then care A is economically more efficient than B.

Quality also has a more subtle dimension relating to the perceptions of consumers of health care. We take demand for health care to be, in part, a function of perceived quality. While, ultimately, this perception may be formed by whether or not the patient's condition improves, perceptions of quality may also be colored by whether or not drugs are prescribed or by how politely the patient is treated by facility staff. We call this definition *perceived quality* of care.

Technical quality and perceived quality are not competing but rather complementary definitions of health care quality. Both are important in understanding health care systems. Technical quality is a supply side concept which involves the medical procedures and their effects on patients' health. Different levels of technical quality imply various degrees of utilization of health care resources and varying levels of patients' health status. Perceived quality is a demand side notion which intervenes in consumers' decisions to seek medical care and to choose a particular provider.

With these definitions of quality, we can go back to the beginning of this section and re-examine our definition of technical efficiency. We will assume that *technical quality of care remains constant along the production possibilities frontier*. That is, not only does quantity of care Q remain constant along the line QQ, but so does technical quality. Thus, not only does any combination of inputs X_a and X_b along the curve permit production of quantity Q of medical care, but also any such combination delivers medical care of constant technical quality, i.e., with the same effect on patients' health status.

Perceived quality of care is not necessarily constant along the frontier QQ, however. For example, while both points 1 and 2 in Exhibit 2-2 permit production of Q visits with the same impact on patients' health, some patients may find point 2 superior, or of better perceived quality. For instance, if X_a is doctor time and X_b is nurse time, some patients may prefer point 2 because it uses more doctor time than point 1, which is more intensive in nurse time.

With the above definitions in mind, we turn to a discussion on efficiency determinants and the effects of provider incentives on efficiency.

3.0 A CONCEPTUAL FRAMEWORK OF PROVIDER INCENTIVES AND GOVERNMENT HEALTH SYSTEM'S PERFORMANCE

The purpose of this chapter is to develop a conceptual framework for analyzing the problem of provider incentives and productive efficiency. The chapter begins with a conceptual discussion that identifies possible sources of technical and economic inefficiency in the production of government health care services. Next, the chapter defines provider incentives distinguishing between monetary and non-monetary incentives. The analysis then suggests how various provider incentives can be used, in theory, to improve technical and economic efficiency. Finally, the possible effects that efficiency incentives may have on equity and quality of care are discussed.

3.1 DETERMINANTS OF TECHNICAL AND ECONOMIC EFFICIENCY

Technical and economic efficiency are affected by both supply and demand side factors. This section is a discussion of efficiency determinants.

3.1.1 Supply and Technical Efficiency

Technical efficiency in government health services is affected by a multitude of factors characterizing the ways and the circumstances under which government health services operate. These factors include: government policies concerning the hiring and firing of personnel; the determination of salaries and other compensation schemes; the regulations and mechanisms governing the acquisition and allocation of other inputs of production (such as capital and pharmaceutical products); the existing norms and resources available defining the technology of production of health services; and many others.

Exhibit 3-1 illustrates how such factors can affect technical efficiency. Suppose, for example, that the amount of services demanded by the population is Q and that facility managers have the ability to choose technically efficient ways of combining inputs X_a and X_b to produce Q units of medical care. In such a case, production of Q could take place at point 1 in the exhibit. Point 1 represents a technically efficient way of producing Q, because if either X_a or X_b were reduced below the levels defined by point 1, then production would necessarily drop below Q.

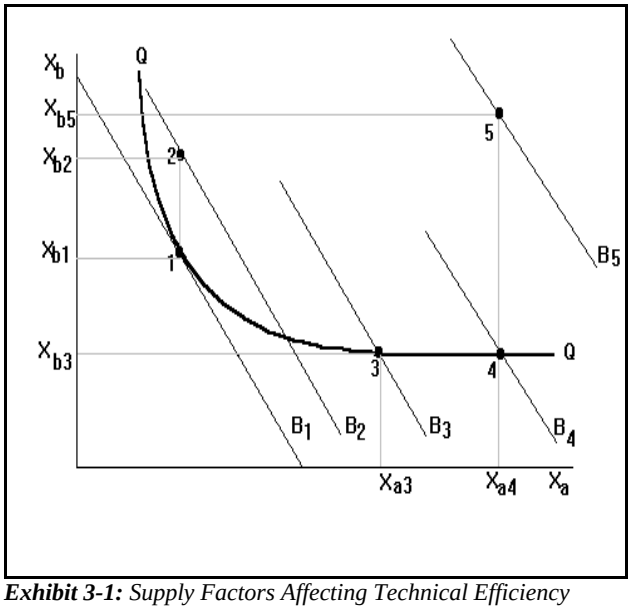

Exhibit 3-1: *Supply Factors Affecting Technical Efficiency*

Suppose, instead, that production of Q occurs at point 2. Point 2 constitutes a technically inefficient way of producing Q because production input X_b could be reduced from X_{b2} to X_{b1} while still producing the same quantity. An example of such a

situation could be one where X_b is nurse labor. If nurses were poorly trained, or if they behaved in a technically inefficient fashion, then production of Q would be technically inefficient. The provision of appropriate training or a change in attitude on the part of the nurses could allow facility managers to reduce the number of nurse hours hired and improve technical efficiency. Their ability to do so will depend on their degree of administrative autonomy. In the presence of administrative autonomy, whether or not the managers will seek to improve technical efficiency will depend upon their incentive structure. The possible types of incentives faced by facility personnel, including managers, and the effect of such incentives on efficiency, are the subject of a subsequent section. Point 2 could also represent a situation where the facility's medical personnel of type X_b choose to spend only a fraction of their contracted time in the facility while illegally devoting the rest to their private practice. If it were possible to enforce full provider compliance with their work hours or if those workers had particular incentives to work their legal hours in the facility, then possibly fewer full-time equivalent employees would be necessary to meet demand. If the excess labor could be laid off or reallocated to other settings, then technical efficiency would improve.

Consider a situation where one of the production inputs (X_b) is available in a limited quantity and equal to X_{b3}, while the amount of X_a input available is X_{a4}. Given the limited substitutability between the two inputs[12], the externally imposed availability of X_a, and the constrained availability of X_b, production takes place at point 4. But production at 4 is technically inefficient: the amount of X_a can be reduced from X_{a4} to X_{a3} while still producing the same quantity of output Q. A typical example corresponding to this situation is a production process that uses some equipment or piece of capital which can be combined with a limited amount of labor. For example, suppose that X_a is surgeon hours and X_b is number of surgery beds available. If a health facility has X_{b3} surgery beds, then any amount of surgeon time available beyond X_{a3} will be superfluous, resulting in technical inefficiency.

Consider, finally, a situation like point 5, where both production inputs are available in large quantities relative to demand, given the technology of production depicted by the production possibilities frontier QQ. Production at 5 is technically inefficient; both X_a and X_b could be reduced to any combination along the curve QQ while still permitting production of Q units of medical care. An example of that situation is one where the level of production resources, such as labor and capital, are exogenously imposed to the facility by central level management of the MOH. If central managers have little knowledge about both demand and the technology of production in the facility, they may overestimate the amount of inputs required to meet demand. If facility managers do not have the authority or the motivation to seek a reduction in the amount of production inputs, then production will be technically inefficient, with the resources being used partially.

[12] To the right of X_{a3} at point 3, substitutability between X_a and X_b is no longer possible; the quantity produced cannot be increased beyond Q even if the quantity of input X_a is increased beyond X_{a3}.

3.1.2 Supply and Economic Efficiency

Most supply-side factors leading to technical inefficiency, such as those described in Section 3.1.1., may also result in economic inefficiency. For example, point 2 in Exhibit 3-1 represents a technical *and* economically inefficient way of producing Q. Production at point 2 requires a total facility budget of B_2. If less X_b were employed, however, a smaller budget of B_1 would suffice to produce the same quantity Q. Purchasing too much X_b (e.g., nurse time) relative to that actually needed to satisfy demand is economically inefficient. Excessive availability of production inputs, given demand and technology, can also lead to economic inefficiency. For example, point 5 in Exhibit 3-1 is technical and economically inefficient: production at point 5 costs B_5, an amount greater than the budget of B_1 required to produce Q at point 1. Constrained availability of one or more production inputs and limited substitutability among inputs can lead to technical and economic inefficiency as well (e.g., too many surgeons for too few surgery beds). For example, point 4, which uses a budget of B_4, is economically inefficient relative to point 3, which uses the smaller budget of B_3.

Not all economically inefficient situations are attributable to technical inefficiency. In many instances, technically efficient production will be economically inefficient. To illustrate this situation, suppose that X_b is available in quantity X_{b5} and consider points 1 and 3 in Exhibit 3-1. While both points have the same technical efficiency, given the relative prices of inputs X_a and X_b, point 3 will be economically inefficient relative to point 1; production of Q according to 3 will cost B_3, while production of the same quantity at point 1 will only cost B_1 ($B_1<B_3$).

So far, two circumstances leading to economic inefficiency have been identified: technical inefficiency (e.g., points 2 and 5 in Exhibit 3-1) and technically efficient production using a mix of inputs that is not cost minimizing (e.g., point 3). There is a third cause of economic inefficiency which arises when the input prices faced by facility managers (for example, personnel wages or pharmaceutical products) depart from social (or shadow) prices. The following discussion centers on the economic inefficiency arising from departures between actual input prices paid for the production inputs, herein referred to as transaction prices, and social prices.

An undervalued production input will likely result in too much use of it relative to what a social economic evaluation would dictate, while an overvalued resource will result in too little use of it. For example, in a country with an overvalued currency, the transaction price of imported inputs, such as pharmaceutical products, will be low relative to the shadow price of those goods. This may result in excessive use of drugs and medical supplies and, thus, ineconomic inefficiency from the viewpoint of society.

Exhibit 3-2 helps to illustrate, through an example, how divergences between transaction and social input prices can result in social economic inefficiency. Suppose that the production of Q curative visits takes place at point 1, the economically efficient way of producing Q when the transaction input prices are W_a and W_b. The associated cost of producing Q is equal to B_1.

Suppose, however, that one of the inputs is undervalued relative to the socially correct price. Assume, for example, that X_b is nurse time and W_b is the undervalued hourly wage of nurses. If the facility manager had to pay the higher, social wage to nurses, the budget constraint line would rotate downwards as depicted by the arrow in the exhibit. In that situation, the budget of B_1 would become insufficient to produce quantity Q. Production of Q at the minimum social cost would occur at point 2 at the higher total cost of B_2. Notice that when the socially appropriate input prices are considered, production at point 1 actually costs society B_3 an amount greater than both B_1 and B_2. Social economic efficiency would increase by moving production from point 1 to point 2 because the social cost of production would drop from B_3 to B_2.

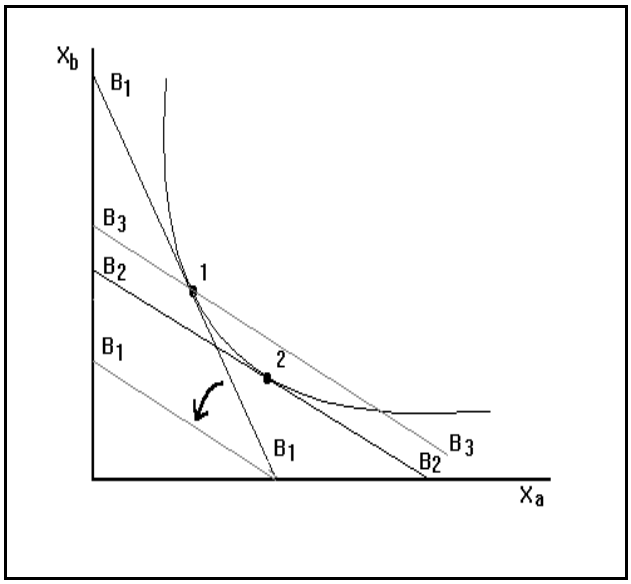

Exhibit 3-2: Social Economic Inefficiency in Production Due to an Undervalued Input

3.1.3 Demand and Technical Efficiency

Though technical efficiency is largely determined by the provider, consumers may be able to affect it by influencing provider behavior through their preferences. For example, certain patients may require that health professionals spend more time in contact with them than is deemed technically necessary or they may demand drugs and exams that are technically not required. This type of behavior can negatively affect technical efficiency. Also, consumer demand is an important factor influencing technical efficiency because it partly determines the degree of utilization of provider resources. Low demand may imply under-utilization of certain resources and thus low technical efficiency, as health facility employees and equipment sit idle. For example, hospitals in rural areas with low population density tend to have low utilization of certain resources, such as specialty care and sophisticated diagnostic equipment characterized by high fixed investment and maintenance costs. Higher demand implies greater utilization of provider resources and thus, other things being equal, greater technical efficiency. Nonetheless, because demand is partly exogenous to the provider, improvements in technical efficiency through provider incentives will be exogenously bound by the level of demand.[13]

[13] Providers have at their disposal several mechanisms to affect demand, such as prices and quality of care. In practice, however, various factors influencing demand, such as illness incidence, competition, and the geographic distribution of the consumers relative to the facility, will impose exogenous limits to the level of demand.

3.1.4. Demand and Economic Efficiency

Patient pressures on providers to operate outside of technical standards, the phenomenon previously mentioned affecting technical efficiency, also affects economic efficiency. In addition, low utilization of resources due to low demand will bring about economic inefficiency.

It is important to distinguish between economic inefficiency arising from excessive, expendable resources being available in relation to demand, and economic inefficiency resulting from low demand relative to the capacity of an indivisible resource of production. Earlier in this section, technical and economic inefficiencies were identified when too many resources were allocated to a facility relative to demand. Because these resources were made up of many individual units (e.g., nurses), it was argued that technical and economic efficiency could be improved if idle resources were laid off. In the case of indivisible resources, however, a reduction in their availability is not possible. For example, certain types of equipment, such as CAT scanners or X-ray machines, are capable of producing large volumes of output. When demand is low, however, the machine becomes under-utilized; the equipment cannot be broken into smaller units, and it is still necessary to incur the full investment and maintenance costs of the asset. A political decision then must be made as to whether or not the service will be provided despite the unavoidable low use and implied low economic efficiency. Certain types of low-demand, often expensive labor, such as some specialized medical services (e.g., ophthalmology, psychiatry) also fit in this category; they can be fully utilized in high demand facilities but are likely to remain under-utilized in low-demand settings, such as facilities in rural areas with low density and scattered populations.

This section has identified several circumstances affecting technical and economic efficiency and has classified them into supply and demand side determinants of efficiency. The causes provided were not intended to be exhaustive; instead, they were drawn as examples from a much larger universe of factors affecting efficiency.

On the provider, or supply, side, various causes of technical inefficiency were mentioned. They included misuse of resources due to poor technical knowledge (e.g., misuse of nurses due to poor nurse training); under-utilization of one or more resources due to constrained availability of one or more production inputs and limited substitutability among inputs (e.g., too many surgeons available relative to a limited number of surgical beds); and excessive availability of production inputs given demand and technology (e.g., too much personnel). Supply-side causes of economic inefficiency included technical inefficiency (e.g., poor nurse training); technically efficient production that is not cost-minimizing; and social economic inefficiency arising from discrepancies between transaction and social prices of production inputs (e.g., undervalued price of nurses leading to overuse).

A demand-side cause of technical and economic inefficiency was the requests and expectations of patients leading providers to depart from technically efficient patterns of production (e.g., to meet the expectations of some patients, providers prescribe drugs even when they are not needed). This behavior also results in economic inefficiency. Another demand side source of

technical and economic inefficiency identified is the level of demand itself. If demand is sufficiently low, certain resources of production such as equipment, buildings, and specialty care may remain under-utilized. Due to their indivisibility, however, under-utilization and economic inefficiency will prevail if a decision is made to offer those services.

A question that has not yet been addressed is the relative importance of technical and economic efficiency: On what type of efficiency should managers wishing to improve productive efficiency focus– technical or economic?

As suggested earlier, economic inefficiency often arises as a result of technical inefficiency. Thus, in many instances, improvements in technical efficiency (reducing the amount of resources used to produce a given quantity of output) will also result in greater economic efficiency (lowering the cost of producing a given quantity of output). An example of this is shown in Exhibit 3-3. Technical and economic efficiency are improved when production goes from point 1 (at cost B_1) to point 2 (at cost B_2<B_1). Unfortunately, greater technical efficiency will not always result in higher economic efficiency. In some instances, a gain in technical efficiency will result in a loss in economic efficiency. For example, production of Q according to point 1 is technically less efficient than according to point 3. Nevertheless, production at point 1 is economically more efficient than at point 3 (because B_1<B_3). This is so because of the relative prices of inputs X_a and X_b.

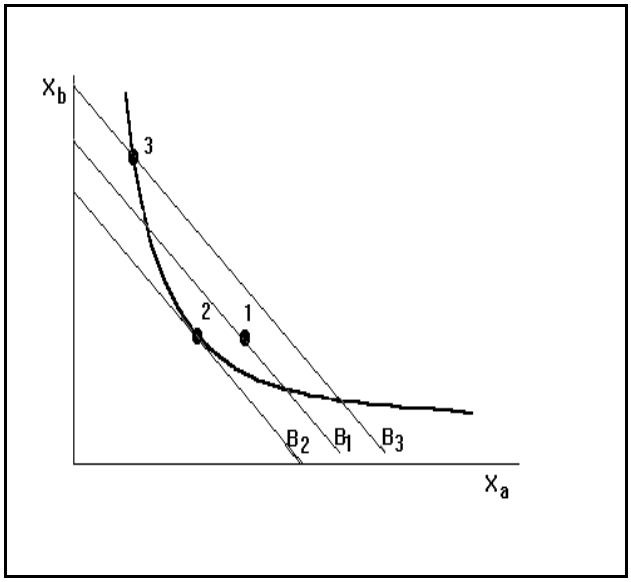

Exhibit 3-3: *Technical Efficiency Vs. Economic Efficiency*

Which of the two production points to choose, point 2 or point 3, will depend on the incentives and constraints of facility managers as well as on quality of care considerations. If managers were free to allocate production resources, and if they had the correct cost-minimizing incentives, they should choose the lowest-cost, economically efficient production, or point 2. One factor which could possibly lead cost-minimizing managers to choose a point along QQ other than point 2 would be the consumers' perceived quality of care. For example, suppose that consumers predominantly preferred point 3 to point 2 because they preferred those services that are intensive in resource X_b. Under those circumstances, if managers attempted to produce at point 2, they could find themselves with a quantity demanded smaller than Q, and with unutilized resources. This discussion illustrates that the problem of deciding how to combine production inputs can become more complex when demand is taken into account and stresses the importance of understanding and considering demand information in the planning process. The ultimate solution to the problem will depend in part on how responsive demand is to alternative input configurations,

as well as on the financial constraints of the facility, the goals of managers, and the mandate of the health system.

A broad range of policy measures can be envisioned to improve productive efficiency of government health services in developing countries. As earlier implied, inefficiency often arises because decisions about the types and levels of production inputs to be allocated to each facility are made by a central bureaucracy, as opposed to facility managers. Centrally made allocations may lead to inefficiency because central decision makers usually have limited information about the conditions—particularly about demand—that prevail in each individual facility. Allocative decisions are often made according to rigid norms (e.g., fixed quantities of beds and personnel per 1,000 population of an arbitrarily defined catchment area) in an effort to reach some rational criteria for resource allocation.

3.2 PROVIDER INCENTIVES

The economic theory of consumer behavior provides a useful framework to study the issue of productive efficiency and incentives. We begin by assuming that a health worker has his or her own *utility* function and faces various constraints. Utility is a measure of a worker's satisfaction in life. The worker's utility is affected by a series of factors, such as the amounts of goods and services the worker and family members can consume, the level of job satisfaction, including the satisfaction drawn from helping others (i.e., patients) through work, and the amount and quality of leisure time. The constraints faced by the worker in seeking utility include income and time. The income constraint states that the worker's household consumption cannot exceed income. A worker's income and the prices of goods and services in the economy indirectly influence the worker's utility because they jointly determine the amounts of goods and services the worker and his or her family can consume. The time constraint establishes that the worker (and family members) have a fixed amount of time to devote to various activities, including work and leisure.

For the purpose of this analysis, it is useful to characterize a health worker's behavior by the degree to which the worker seeks to achieve efficiency at work. The worker behaves with the aim of maximizing utility subject to constraints. Without attention by employers to specific incentives for the workers' performance, this behavior may—and most commonly will—result in levels of efficiency (as well as equity and quality) which depart from what the employer would like the worker to achieve. Incentives are factors or circumstances which affect worker behavior. Incentives can intentionally be established by policymakers to induce particular types of behavior, although not all provider incentives are established by, or under the control of, policymakers. Provider incentives constitute a set of policy instruments which can be used to influence the worker's behavior with the aim of narrowing the gap between the worker's performance and the desired performance.

From a policy perspective, incentives can more precisely be defined as a series of mechanisms which link a worker's performance with the monetary and non-monetary rewards the worker receives for performance. There is an important distinction between the determinants of a worker's utility and incentives, which

may be clarified through an example. While a worker's remuneration influences his or her utility, it will not necessarily affect behavior unless the remuneration is attached to the worker's performance through particular conditions which provide specific incentives. For example, a worker's salary may be doubled, but this will not necessarily affect behavior.[14] In contrast, a worker's remuneration may remain the same, yet it can affect the worker's behavior. This can be done by somehow linking the remuneration level to the worker's performance. For instance, if the worker's pay is proportional to the volume of service delivered, this will provide an incentive to increase output since income will increase and, thus, the worker's utility. If his pay comes in the form of a fixed salary, irrespective of performance (as is the case in many developing country public health systems), one should not be surprised if performance is poor.

Provider incentives can seek to reward desired worker performance by providing the worker with goods or services that directly enhance his utility or do so indirectly by modifying the worker's income or time constraints. Examples of the latter include vacation time and monetary incentives which attach the worker's remuneration, or the price of certain goods and services (e.g., subsidized health care and housing for the worker's family), to performance.

Provider incentives can also be classified as monetary and non-monetary and targeted to have an effect on one or more determinants of provider behavior. Monetary determinants include the take-home income of the employee, which may consist of a fixed salary, performance bonuses, and additional income that may be tied to patient payments or to third-party reimbursement. Non-monetary determinants include the quality of the work environment (e.g., friendliness of colleagues, characteristics of the premises), the employee's technical skills, his or her job security, non-monetary job-related benefits, community recognition, and other non-economic factors which influence the worker's behavior.

3.3 EFFICIENCY INCENTIVES

As noted in Section 3.1, a chief cause of inefficiency is the fact that resource allocation decisions concerning production inputs are often made by central authorities instead of facility managers. Poor personnel training, either in technical matters or in management techniques, can also result in low efficiency. The lack of production and managerial resources can hamper technical and economic efficiency, as well. Distortionary pricing can also be a source of social economic inefficiency.

Even when there is some room available for the employees to act more efficiently, however, they may fail to do so due to the lack of both incentives that reward efficient actions and penalties that discourage the opposite.

[14] An example of this is reported in Section 5.0, "Review of Developing Country Experience with Provider Incentive Schemes," for the case of Cote d'Ivoire.

Several policies can therefore be envisioned to achieve greater efficiency in government health services. They include, among others:

(1) Promoting greater decentralization of decision making;

(2) Improving personnel training in medical and management techniques;

(3) Improving resource availability;

(4) Removing price distortions; and

(5) Designing appropriate personnel incentive systems, including alternative health care financing schemes.

The establishment of provider incentives is thus viewed not as the unique solution to the efficiency problem, but rather as an integral part of a broader reform package. The success of efficiency boosting incentive programs will depend largely on the degree to which policies such as (1) through (4) are adopted. At the same time, the success of policies (1) through (4) will be partly contingent on the adoption of appropriate provider incentive systems.

A central hypothesis of this paper is that efficiency can be improved through appropriate personnel incentives. These incentives can be monetary or non-monetary. Possible monetary incentives that can be adopted are discussed first, the adoption non-monetary incentives is discussed next.

3.3.1 Monetary Incentives

Traditionally, government health workers in developing countries obtain their remuneration in the form of a fixed weekly or monthly payment. This payment, or salary, is called *fixed* because, while it may vary over time reflecting raises or inflation adjustments, it is generally fixed relative to the performance of the worker. In most of the developing world, government jobs are also more secure than equivalent private sector jobs. Security, in this context, means that government employees are less accountable for their performance, and thus less likely to lose their jobs due to poor performance. (In addition, they are less likely to be rewarded as a result of good performance.) Job security is thus intimately tied to employee behavior and accountability. Other things being equal, secure jobs are likely to result in poor performance.

In addition to constant salaries and secure jobs, there is a third circumstance prevalent in developing country health systems which also has detrimental effects on health worker efficiency. It is well known that, in most of the world where there is a lucrative private market for health workers, government health employees also practice in the private sector. Examples of such instances are documented for the Dominican Republic (Lewis et al., 1990).[15] While in many cases this behavior is considered illegal, it is simply a reflection of rational behavior on the part of the government employees seeking

[15] This issue is discussed at length in another HFS Phase 1 paper by Ellis and Chawla (1992).

to maximize their utility. The prospect of sufficiently high, additional income from private practice will lure many public servants to work privately, sometimes using government infrastructure and resources.[16] If job security is high, this behavior will be pervasive; it some instances, it will mean that government services will likely be inefficient and of low quality. Inefficiency will result from under-utilization of labor, as some workers spend only a fraction of their contracted time in government facilities, while devoting the rest to private practice.

The combination of constant salaries, secure jobs, and money-earning opportunities in the private market is likely to have pervasive effects on productive efficiency, according to our utility maximizing framework of Section 3.2. Utility-seeking health workers, facing few to no risks from poor performance, will seek to boost their income by working in the private sector while minimizing the level of effort in their government jobs.

Three types of efficiency incentives can be envisioned from the above discussion: removing tenure or job security; linking workers' remuneration to the level of efficiency achieved in production; and providing, if necessary, supplemental income to health workers to make their practice in government facilities as lucrative as private practice.

Removing tenure of government jobs is an important yet difficult-to-achieve policy. It is one that is hard to implement in one sector alone, such as health, without changing the entire civil service code. Nevertheless, as long as public workers lack the threat, fundamental in the private sector, of losing their jobs in the face of poor performance, any efficiency reform that relies on incentives will have only limited effects.

Linking the workers' remuneration to their productive efficiency is, to the best of the authors' knowledge, a policy not yet adopted in the developing world. In contrast, as is shown in the following chapter, that system has been implemented in some health care organizations in the U.S. A possible mechanism to make such a system operational is one where health facilities are assigned by the government a fixed budget, or subsidy. Health workers are allowed to determine how those funds are used to purchase production inputs. Any funds remaining at the end of the budgetary period must be declared to the government and can either be distributed in full among the workers as a performance bonus or split between the government and the workers. The resulting patterns of resource utilization can be used by the government to produce more efficient budgets in subsequent periods, while saving money and still assigning larger-than-necessary subsidies to provide the workers with the monetary incentive to keep efficiency high.

Supplementing the income of health workers is an increasingly popular practice in developing country health systems. In recent years, cost recovery for health care and drug revolving funds have become common policies in government

[16] As long as the marginal income from outside work is greater than the marginal value of time, government workers with high job security will seek outside wages. Security of government jobs is the important factor behind moonlighting, not how high government pay is, as is commonly thought.

health systems in many African countries (usually in the form of the Bamako Initiative), and in much of the developing world. Cost recovery often provides government health workers with a mechanism for supplementing their often meager salaries.[17] This constitutes an incentive for the workers—especially for physicians—to devote more of their time to their government practices, thereby improving technical and economic efficiency.[18]

Cost recovery not only provides a financial incentive for the health personnel, but also allows health facilities to generate income with which to pay for needed production inputs, chiefly pharmaceutical products. That permits health facilities to improve productive efficiency since the lack, or the limited availability, of production inputs was identified as a cause of inefficiency in Section 3.1.

There is a third reason why cost recovery, in the form of a Bamako Initiative-like drug revolving fund, can improve productive efficiency by providing health workers with particular incentives. When health services, including drugs, are provided free of charge to the patients and allocated to the facility by the government unconditionally, inefficient prescription practices are likely to arise. For example, poor diagnostic and prescription practices are a serious problem in government health facilities in The Gambia. There, *poly-pharmacy,* or the over-prescription of often expensive, unnecessary drugs to patients, leads to a major waste and to a health threat. Since fixed-sized, monthly shipments of drugs are periodically sent by the MOH Central Stores to government facilities, health workers see little gain in prescribing drugs in a medically responsible fashion. The adoption of adequate prescription practices usually requires additional effort by the workers in the form of further physical examinations and laboratory tests. Health workers see little benefit in incurring such an additional effort; to the contrary, they believe that poly-pharmacy increases the probability of curing patients with limited effort and reduces the patient load by decreasing patient returns.

Ministry health facilities in The Gambia have adopted a fee-for-service system for curative patients. This fee is fixed for all patients irrespective of the amount of drugs they are given. In order to contain pharmaceutical costs and to reduce poly-pharmacy, the MOH is considering a proposal to link facility cost recovery revenue to the amount of pharmaceuticals shipped to the facility.[19] This measure will most likely be effective in reducing drug consumption and poly-pharmacy. At the same time, it will provide facility personnel with an incentive to enforce fees to be able to maintain a constant volume of funds with which to replenish their drug revolving fund.

---

[17] For an example of how cost recovery has helped supplement health workers' salaries, see Bitran et al., 1987.

[18] Efficiency will be improved only if demand is sufficiently high, or, equivalently, if labor levels are in accordance with demand. If demand is low relative to the full-time equivalent medical personnel, then efficiency will be low whether the personnel spend part of their time in the facility doing nothing or spend that time working in their private practices.

[19] Tilney, et al., 1992.

Splitting cost savings between the government and health workers, a measure discussed earlier, constitutes one of many conceivable incentive measures which involve changing the way government health services are financed. In fact, there are many alternative ways in which the government can finance governmental health providers and each has its own implications on efficiency. The industrialized world offers many examples of alternative modes of government financing of health care. So does the private sector in most countries. Inevitably, however, many of those financing systems would imply a drastic change in the way government health systems operate. Financing schemes such as the U.S. DRGs or the Canadian government reimbursement system differ greatly from the present subsidy-combined-with-cost-recovery system of many developing countries. A switch in financing mode would undoubtedly constitute a major departure from the present provider incentive structures tied to cost recovery and government lump-sum subsidization. Because we consider such a financing reform unlikely in the near future in most developing countries, its likely effects on provider incentives and efficiency is not discussed. The following chapter, however, does discuss the incentive and efficiency implications of such financing schemes in the context of the industrialized world.

3.3.2 Non-Monetary Incentives

Non-monetary, or in-kind, incentives can take a wide variety of forms. Examples of non-monetary incentives include: vacations, increased management authority, housing, child care, travel, improved work environment, and educational benefits. The implications of these various types of non-monetary incentives do not depend on the specific form of benefit. From the provider's perspective, the only difference in this class of incentives is that the precise pay-off is not in cash income. From the government's perspective, however, where the aggregate cost of implementing the incentive program might be a critical concern, the distinction between monetary and non-monetary incentives can be important. In that regard, the composition of public resources available to underwrite the program would have a large effect on the choice of incentive type.

The critical factor in determining the effects of these non-monetary incentives is the condition under which these benefits are offered. For example, these incentives could be tied to cost containment or efficiency enhancement at the facility. The conditions for receiving rewards can take various specific forms. For instance, they can depend on meeting specific performance targets, or they can be given in increasing amounts as a function of the extent to which specific targets are exceeded.

Non-monetary benefits may be particularly useful devices in situations where providers' incomes are fixed by civil service standards, or where insufficient cash exists to finance direct income supplements. More specific arrangements are also possible. For instance, an effort to increase equity in public health care might translate into efforts to increase the number of physicians in rural areas. Non-monetary incentives, such as subsidized housing or the promise of specialty medical training, might be used to entice physicians to serve for some fixed period in rural locations.

Non-monetary incentives can also contribute, not to the real incomes of staff, but to improving the quality of the environment in which they work. For example, staff could be permitted to spend some portion of revenue or profits to upgrade their clinical equipment or physical plant or they could be provided with additional training. This strategy would address the notion that work conditions affect staff morale.

3.4 EFFICIENCY INCENTIVES, QUALITY, AND EQUITY

The preceding section was a discussion of possible monetary and non-monetary incentives that can be adopted to reach greater productive efficiency in government health services. Missing from the discussion was the likely effects of those incentives on two important dimensions of any health system, namely equity and quality of care. This aspect is discussed in this final section of this chapter.

3.4.1 Relationship Between Efficiency and Equity

Economists often consider the conflicts between efficiency and equity. In the present context, for example, if efficiency were defined as the number of children immunized per dollar spent, it would be more efficient to vaccinate children living near health facilities because the transportation costs incurred to reach nearby children are lower than for distant children. Yet that approach would create an inequity for children living far from the facility, since they would be much less likely to be vaccinated. Correcting this inequity by attempting to vaccinate all children (regardless of location) with equal probability would most likely come at some cost to the program's efficiency. Incentives seeking to improve the efficiency of vaccination activities would have to be devised carefully to avoid the possible tendency of health workers to behave inequitably.

Efficiency and equity need not always conflict. In particular, efficiency and equity can both improve at the same time. Consider, for example, a situation where full-time health workers show up at the facility for only a few hours each day, early in the morning, as is the case in some government health centers in the Dominican Republic. This is a very inefficient situation since these workers are paid full-time salaries; thus, the cost per visit provided is very high. But this is also inequitable because those who are able to obtain care are those who can arrive at the facility early enough (most likely those living near the facility) to catch the provider. The gain in efficiency that would be achieved by requiring, or inducing through incentives, doctors to work more hours at the facility would also enhance equity by allowing other population groups to obtain care as well.

There may also be a set of efficiency-enhancing activities that are neutral with respect to equity. In those cases, one might be able to increase efficiency without affecting anyone's accessibility to constant-quality care. In general, the possibility for increases in either efficiency or equity which are either neutral or positive with respect to the other requires that there be ample room to improve along both dimensions. An example of this might be efficiency gains to keep up with population growth.

3.4.2 Relationship between Efficiency and Quality

There is also concern that greater efficiency may often result in lower quality health care.[20] In this regard, one must distinguish clearly between efficiency gains and cost reduction. More specifically, one must distinguish between unit cost reduction (i.e., efficiency gains) and aggregate cost reduction. The notion of a unit cost reduction implies that the unit itself remains unchanged, which implies that the quality of that unit of care does not change.

The distinction pertains to the source of the efficiency gain. Wyszewianski, et al. suggest that quality may actually improve when efficiency gains come through the elimination of unnecessary services. Since nearly every procedure entails some degree of risk to the patient, they argue that the elimination of an unnecessary procedure must result in net benefits to the patient.

On the other hand, they concede that the quality of care diminishes when useful services are eliminated. Yet, by the definitions of efficiency presented above, the elimination of needed services would not constitute an efficiency gain as opposed to simply a reduction in aggregate costs. For instance, it is not more efficient to release patients from the hospital prematurely; while it would reduce the aggregate cost of the hospital stay, it would result in a drop in technical quality.

Various types of efficiency incentives were discussed in the previous sections. Unless special attention is devoted to the way those incentives are provided, they may have detrimental effects on equity and quality of care. For example, the incentive that allows health workers to retain all or part of their cost savings may have pervasive effects on quality of care if needed procedures and services are eliminated in an effort to reduce costs. Thus, such an incentive would have to be accompanied by measures to avoid quality reductions. A possible measure would be government monitoring of provider practices and contrasting of those with standards of treatment.

In some instances, market conditions may help minimize the extent to which quality suffers as a result of the attempts of providers to reduce costs. A common example is that of a health care market where government facilities operate with a user fee system and compete with private providers. Quality reductions, in the form of fewer pharmaceutical products available or less personnel time with the patient, may lead patients to seek care elsewhere. If facility personnel draw a financial incentive from cost recovery revenue, they will self-limit their tendency to reduce quality beyond the level under which their financial incentive from cost recovery begins to deteriorate.

In the light of the popularity of cost recovery in the form of user fees, we conclude this chapter with a brief discussion of its possible effects on efficiency, quality of care, and equity.

---

[20] See, for example, Wyszewianski, et al., 1987.

Consider the common case of a facility where the medical staff receive a fixed salary from the government, irrespective of performance, and a variable income supplement drawn from cost recovery from fee-for-service patient payment. More specifically, suppose that each employee gets a fixed percentage of the constant price of, for example, each patient visit. Assume also that there is some upward flexibility in the government budget for drugs and other production supplies. Under these circumstances, it is clear that, as long as the amount of money each employee gets per visit is worth more to him than the effort required to perform a visit, then he will seek to maximize the quantity of visits produced. This behavior will likely be inefficient and inequitable. Depending on market competitiveness, this behavior may also result in low quality.

Inefficiency will come from two sources. First, the provider may substitute supplies for his own labor to reduce the effort of each additional visit. This will increase marginal and average cost. Second, in his effort to increase production, the employee may require patients to return for additional visits, which may bring about little gains in health status to the patient. There will be some demand-side constraints on this behavior, however, because patients have to pay both out-of-pocket for each visit and to incur time costs. In addition, to the extent that patients perceive the drop in quality, this will also depress their demand.

Inequity may result from the employees' reluctance to treat patients with limited ability to pay. With regard to quality of care, the employee's interest in producing as many visits as possible discussed earlier may in fact hurt the patient's health. The provider's desire to reduce his effort per visit is also likely to result in poor quality care. If competition in the market is low, patients will have few options but the lower quality care.

Consider a modified example with widely different implications. Suppose that the government no longer pays for drugs and other medical supplies, and cost recovery revenue, net of provider pay, is used to pay for those items. Suppose also that all the cost recovery revenue remains at the facility. This is an increasingly common situation in the developing world, especially in Africa. Faced with a budget constraint, the workers will operate differently. In fact, given that the price per visit is fixed by the government, they will be inclined to treat low-cost problems, to reduce treatment costs, or both. The first behavior implies that the workers may attempt to deny or limit care to patients with complex problems. This is not necessarily inequitable if health problems affect all population groups equally. The second behavior, cost minimization, would result in greater technical and economic efficiency if quality of care were maintained. However, in the absence of quality standards and assuming limited patient ability to distinguish and demand quality care, efforts to reduce costs may result in poor quality care. Compared with the previous example, this case would promote greater efficiency, though likely with more negative consequences for quality of care as well as with reduced access to care.

A third, more complex example is one where the provider is allowed to fix his own prices. The reader can anticipate the effects of this on efficiency, equity, and quality. In sum, the mechanism of provider compensation is expected to have an important effect on the three dimensions of provider performance. Legal, practical, or other types of limitations in the types of compensation

mechanisms that are possible in government health care systems generally imply that most feasible systems have both positive and negative implications for performance. This calls for efforts to reform compensation systems or methods of targeting incentives.

The following section is a discussion of provider incentive systems in the industrialized world and their empirical effects on provider performance. A similar review is presented in Section 5.0 for the developing world.

4.0 OVERVIEW OF DEVELOPED COUNTRY LITERATURE AND EXPERIENCE

The literature on health care provider incentives in developed countries falls broadly within two categories: physician incentives in ambulatory care settings and hospital incentives for cost containment. This section briefly reviews that literature and evaluates its relevance for developing countries.

4.1 PHYSICIAN INCENTIVES IN AMBULATORY CARE SETTINGS IN THE U.S.

The literature on physician incentives in ambulatory care settings revolves largely around the implications of alternative reimbursement structures in health maintenance organizations (HMOs) and individual practice settings. The basic reimbursement modes include fee-for-service, salary, and capitation. Recent literature examines the particular incentive effects of alternative reimbursement and institutional structures. Yet these questions could not be addressed had the earlier literature not first established the basic point (which was not taken for granted) that physicians are indeed responsive to financial incentives.

Under third-party fee-for-service reimbursement, the physician's income increases with the number of services provided.[21] This system is widely thought to provide physicians the incentive to run up the bill by delivering more care than is necessary.[22] This was the prevailing system, for example, in the U.S. Medicare program during the 1960s. The resulting pressure to overuse medical care later resulted in significant revisions in the Medicare reimbursement system and penalties for physicians who failed to meet the new standards for practice.

The rise of HMOs in recent years has accompanied a growing concern with health services cost containment. The HMO model has made it more common than was previously the case for physician compensation to be based not on fee-for-service, but on a salaried or capitation basis. In fact, it is slightly misleading to speak of the "HMO model", since HMOs are commonly structured in one of four different ways:

1) the *staff model*, in which physicians are hired directly as staff;
2) the *group model*, in which an HMO contracts with independent groups of physicians to provide services;
3) the *network model*, which is similar to the group model, but involves several groups; and
4) the *individual practice association*, in which the HMO contracts with physicians in individual practice.

The reimbursement structures listed above cut across these specific HMO models. Under a capitation system, physicians are paid a flat rate per enrollee-month. In other words, the physician receives a fixed payment for each HMO member who designates him or her as the primary care physician (PCP), in return

[21] It is assumed that patients either do not pay out-of-pocket or their payment (deductible or copayment) is small enough so that demand is price inelastic. Under those circumstances, physicians can increase the number of services provided with little concern about losing customers.

[22] See, for example, Egdahl and Taft (1986), and Relman (1988).

for the promise of comprehensive care. Compensation in this system does not increase with the quantity of care provided. To the contrary, the physician or the HMO faces potentially large down-side risks from having to provide expensive care at that fixed price. This system thus creates an incentive for physicians to minimize the number of patient visits. While there is no incentive for the physician to save non-physician costs incurred in treating patients (these costs are paid by the HMO), there is also no incentive to overprescribe care (other than to protect against the risk of malpractice suits), since the physician's income is not a function of quantity of care provided.

Moving from fee-for-service to a strict capitation system improves economic efficiency, while having ambiguous consequences for equity and quality of care. The improvement in economic efficiency comes from eliminating the incentive to overprescribe. In terms of equity, while fee-for-service creates an incentive to attract patients (i.e., to promote use), capitation provides an incentive to attract enrollees but discourage actual use. To the extent that physicians do not discriminate among patients in their efforts to dissuade them from using physician services, capitation does not have clear equity implications. Capitation, however, is likely to have negative consequences on access to care for the reasons provided earlier. Capitation would create conflicting incentives with respect to quality, since there are incentives both to discourage patient visits and maximize enrollment (which depends, in part, on a physician's reputation for providing quality care).

The salient point is that capitation and salary reimbursement structures provide a mechanism through which HMOs can create incentives to change physician behavior by putting their compensation at risk. For instance, it is common for individual practice associations to withhold 15 to 30 percent of the physicians' fees until the end of the year, at which time it is distributed to physicians who have met pre-specified goals for performance.[23] Incentives may take the form of either rewards or penalties, and are generally awarded based on some notion of cost containment. A great deal of the developed country literature on physician incentives thus addresses the question of whether putting physician compensation at risk creates a conflict of interest vis-a-vis the quality of health care.

Several examples illustrate this concern. In many HMOs, the primary care physician serves as the "gate-keeper" with regard to costs incurred for specialty care, consultants, and hospitalization. Thus, to the extent that PCP compensation is tied to cost minimization, there is an incentive not to use specialists and to keep patients out of the hospital.

A more specific example is that of Blue Cross-Blue Shield of North Carolina, which in 1982 began offering bonuses of up to 25 percent to surgeons who carry out certain procedures in their offices rather than in the hospital.[24] More recently, Blue Shield of Massachusetts instituted a pilot program under which obstetricians are paid more if their patients have shorter-than-average lengths of hospital stays.

---

[23] Egdahl and Taft (1986).

[24] *Ibid.*

As discussed above, the elimination of unnecessary procedures would reflect in improvement in the quality of care. Yet concern arises from the lack of guarantee that beneficial procedures will not be foregone when doing so increases physician compensation.

Indeed, recent statistical analysis has found evidence that the use of capitation or salaries is associated with a lower rate of hospitalization than is found under the fee-for-service system.[25] The same study found that hospitalization rates were lower for physicians in for-profit and group model HMOs, and that placing physicians at personal financial risk was associated with fewer outpatient visits per enrollee and a higher probability of the HMO breaking even.[26] Unfortunately, this study stopped short of drawing conclusions regarding the effect of these findings on the quality of care.

Another study examined the effects of an incentive program introduced by a chain of for-profit ambulatory care centers.[27] In this case, salaried physicians were provided financial incentives to increase revenues. Instead of receiving a flat hourly compensation, physicians in this organization were to receive either the flat rate or a percentage of the gross monthly charges they generated, whichever was higher.

Comparing practices before and after the initiation of this incentive program, the study found that physicians increased the number of laboratory tests performed per patient visit by 23 percent and the number of x-ray films per visit by 16 percent.[28] In addition, this study found that total charges per month grew 20 percent (in part, due to a 12 percent increase in the average number of patient visits per month), and that the wages of the physicians who regularly earned the bonus rose by 19 percent.

More complex combinations of these various incentive structures exist. Many of the variants have to do with the share and disposition of physicians' income withheld and put at risk against some performance standard. Many of the new Medicaid programs, for example, use a partial capitation system (in which consumers are promised only partial coverage in return for a fixed prepayment).[29] Yet, the present discussion serves to illustrate the basic issues relating to physician incentives and HMOs in developed countries.

---

[25] Hillman, Pauly, and Kerstein (1989).

[26] *Ibid.*

[27] Hemenway, Killen, et al. (1990).

[28] *Ibid.*

[29] See Welch (1990) for a review of such programs.

Much of the remaining developed country literature on provider incentives reviewed concentrates on hospital-level issues.

4.2 HOSPITAL INCENTIVE LITERATURE IN THE U.S.

The primary question that arises in the hospital incentive literature in industrialized countries is: does hospital productive efficiency depend on whether the hospital operates for profit or not for profit? A related and much debated question in the economic literature is: what do hospitals maximize? Other segments of the literature deal with cost containment in Medicaid and Medicare, and the details of complicated reimbursement and pricing structures (such as diagnostic related groups, or DRGs).

A theoretical argument that non-profit hospitals must be relatively inefficient derives from the theory of property rights. The notion is that, since no individual can augment his or her income by minimizing costs, the non-profit hospital is inherently inefficient.[30] Baird further contends that one cannot assume that non-profit hospitals minimize costs, and that models based on this assumption are invalid.

Clarkson extends this view, arguing that the owners of for-profit hospitals have exclusive rights to the flow of cost-savings arising from the reduction of hospital care, and that the owners maximize these benefits by hiring managers whose incomes are made a positive function of profits.[31] In contrast, the link between income and managerial performance in non-profit hospitals is much weaker, and Clarkson asserts that this leads to inefficient behavior.

Empirical studies of for-profit and non-profit hospitals have tended not to support the property rights theory. One study, for example, found that while for-profit hospitals charged more and were more profitable than non-profit hospitals, there were no differences in productive efficiency attributable to ownership of the hospital.[32] That study examined several alternative proxies for productivity, including total asset turnover ratios (gross revenue divided by total assets), case flow (admissions divided by beds), average lengths of stay, occupancy rates, and full-time equivalent staff. In each case, ownership structure failed to explain measurable productivity differences (holding constant a range of other variables). Another study, concentrating on lengths of stay as a measure of efficiency, also failed to find significant differences between for-profit and non-profit hospitals (controlling for case mix and other variables).[33]

---

[30] This argument, first presented by Baird (1971), is summarized in Register and Bruning (1983).

[31] Clarkson (1972), p. 364.

[32] Renn, Schramm, Watt, and Derson (1985).

[33] Freund, Shachtman, Ruffin, and Quade (1985).

This issue revolves around the question of what objectives hospitals maximize. Barnum and Kutzin provide a partial list of alternative behavioral models of hospitals, which includes:

- Maximizing output (patient admissions), given a fixed budget;
- Maximizing some function of output and quality of care (assuming a trade-off between the two);
- Minimizing cost, given exogenous demand for admissions;
- Maximizing some function of profit and output;
- Maximizing institutional prestige, which is a function of hospital size, facilities, and the prestige of associated physicians; and
- Satisfying models of behavior, where managers and staff only hope to achieve some level of output and quality within a fixed budget that will satisfy their own and higher-level managers' expectations.[34]

Each of these objective functions implies a different set of managerial incentives. Yet Barnum and Kutzin (citing Pauly) point out that it is virtually impossible to distinguish econometrically between most of these alternative objective functions.[35]

While some analysts treat the hospital as a monolithic whole, other seek insights by disaggregating hospitals' internal actors. Harris, for example, concludes that the hospital as it operates in the U.S., where patients or their insurers pay fees to both the hospital and the doctors, is actually two separate firms: the medical staff (which creates demand) and the administration (which supplies hospital services to meet physicians' demand).[36] He finds that each "firm" has its own managers, objectives, pricing strategies, and constraints. Harris uses this perspective to analyze internal resource allocation decisions and the shortcomings of U.S. regulatory policy. As Barnum and Kutzin point out, however, this perspective is less relevant for hospitals in many developing countries, where the administrative and medical staffs are more likely to coincide.[37]

Indeed, it is important to ask generally how relevant the developed country experience is for developing country health sectors. The following section addresses the special incentive issues facing health sectors in developing countries.

---

[34] Barnum and Kutzin (1990).

[35] *Ibid.*

[36] Harris (1977).

[37] Barnum and Kutzin, *op. cit.*

4.3 SYSTEM-WIDE INCENTIVES IN DEVELOPED COUNTRIES

A wide variety of actual practice exists in developed countries, often combining various elements of alternative incentive structures. A recent survey of physician compensation schemes has found that the level of compensation is determined through negotiation between the physicians and the payer in most industrial countries.[38] In particular, the most common structure involves such negotiation in the context of a fee-for-service system. Variations on this structure, where the payer is either the individual or a third party, exist in Belgium, France, Switzerland, Norway, other countries in northern Europe and North America, and in New Zealand. A similar, though less common, arrangement exists in Quebec, where the object of negotiation is physician salary.

Other countries have favored price setting by administrative decision. Although the survey encountered only rare examples of administrative pricing combined with fee-for-service, administrative pricing was the most common approach when the object of compensation was salary. This approach is found in the former Soviet Union, across Eastern Europe, Spain, and for specialists in Great Britain, Italy, and Ireland.[39]

The only compensation structure found to be compatible with the determination of payments by physicians themselves was fee-for-service. Regardless of whether payment is by individuals or third parties, physician-determined fees-for-service are found almost exclusively in the United States. Among the various combinations of such structures encountered in this survey, the greatest combination of modalities was found to co-exist in the United States. Only in the former Soviet Union and Eastern European countries was there only one official approach to determining physician compensation.

More detailed case studies exist for several countries. For instance, issues of provider incentives and productive efficiency have received great attention in recent efforts to reform the Dutch health care system.[40] In large part, these reform efforts have been the result of local experimentation.

General practitioners in the Netherlands serve as "gatekeepers" to the health care system by controlling referrals to specialists and to hospitals. Depending on the patients' income, the general practitioners are paid on either a fee-for-service basis or on a capitation basis.[41] Specialists are paid on a fee-for-service basis. Under this system, both types of rate structures are subject to strict government regulation and are determined by negotiation between physicians and the third party payer associations.

---

[38] Contandriopoulos, Champagne, and Baris (1993)

[39] *Ibid.*

[40] Kirkman-Liff and van de Ven (1989)

[41] The former tends to apply to the treatment of wealthier patients, who generally have private insurance; the latter system pertains to the treatment of less affluent patients, who generally belong to local semi-public insurance funds.

The previously existing system in the Netherlands conveyed numerous perverse incentives.[42] For instance, the capitation system rewarded general practitioners for referring patients to specialists, even although the general practitioners themselves ight have provided treatment in many cases. The fee-for-service system for specialists created incentives for unnecessary and unnecessarily expensive treatments. The hospital financing system made it more attractive to admit patients for three days than to perform lower-cost day-surgery. In addition, the local semi-public insurance funds received full coverage of losses from a national fund; more efficient funds accordingly received less revenue from the general fund, thus undermining incentives to improve efficiency. Rapidly rising health care costs under this system have necessitated widespread reform.

Efforts to improve efficiency through revised economic incentives have centered on the general practitioners' role as gatekeepers. In particular, the goal has been to reduce the number of referrals to specialists and hospitalization. One local insurance fund in the Netherlands offered bonus payments in addition to the standard capitation payment to general practitioners who achieved pre-specified cost containment goals and penalized those who failed to meet their goals. That local experiment succeeded in reducing utilization rates by 10 percent, although no measure of changes in the quality of care were recorded. Other local experiments in the Netherlands tested a wide range of reforms. One fund, for instance, introduced a combination of capitation and extra fee-for-service for general practitioners. The extra fees, though lower than the levels offered specialists, discourage referrals for certain minor procedures. Another fund adopted a similar system, with the addition of a bonus/penalty system involving the costs of prescription drugs and physical therapy.

The English National Health Service (NHS) provides a contrasting example.[43] Physicians under the NHS, at least as structured in the late 1980s, had little incentive to be efficient. The State met all costs of care, and specialists were appointed for life.

Studies have identified various sources of inefficiency in the English NHS. One source pertains to the compensation structure for general practitioners (GPs) working in the NHS. While GPs are paid on a capitation basis, this source on average comprises less than half of the GPs' total remuneration from the NHS. Other components include basic practice allowances, bonuses for working in certain areas, payments for training apprentices, and additional fees for providing certain services. The GPs time and output were not closely monitored, and physician compensation was not performance-related.

Additional sources of inefficiency pertained to the financial relationship between the GP service and the Hospital and Community Health Services branch of the NHS, which, unlike the GP service, was cash-constrained. For instance, diagnostic tests, such as X-rays and blood tests, are prescribed by GPs but

---

[42] See Kirkman-Liff and van de Ven for greater detail and further references.

[43] **See Maynard (1987) for details.**

provided free of charge by the hospital service. Similarly, GPs can use hospital specialists to reduce their own work load, shifting the costs to the cash-limited hospital service. There are no incentives for the GPs to control costs or to limit referrals.

The Government's efforts to reform this system, in contrast to reforms in the Netherlands, were heavily centralized, with little scope for small-scale local experimentation or review. Moreover, system-wide budgetary reforms left physicians' contracts unchanged. As before, physicians' contracts were held by regional authorities rather than the facilities where the physicians actually worked. Appointments were still for life, physicians were still not held financially accountable for their practices, and local facilities had no authority to discipline poor performers.

The Polish National Health Service provides a more extreme example.[44] For the first forty years after its establishment in the 1940s, the Polish NHS operated on the ideal that free health care was everyone's moral right. The health system was seriously under-funded during this period, and compensation of health care professionals was lower than that of other professionals. These factors required the rationing of health care, which continued to be provided essentially free of charge.

NHS physicians were traditionally paid on a salary basis, independent of the both the amount and the quality of their work in the system. This created a strong incentive for physicians to minimize their patient contacts within the NHS, and to divert their time to outside work. Approximately 80 percent of NHS physicians were found regularly to work more than one job.[45] In addition, health care institutions were given little autonomy in financial planning. Provincial authorities allocated annual budgets to local institutions. The incentive for these institutions was to spend their entire allocation, for fear of having it reduced the following year.

Beginning in the 1980s, the Polish Government undertook a series of reforms designed to introduce economic incentives for better performance into the NHS. The major innovation came in 1985, with the introduction of a bonus system for health care workers. Awarded on the basis of the amount and quality of work performed, these bonuses could constitute over one-third of the recipient's salary. There were also various experiments in which physicians were allowed to compete for patients under either a fee-for-service or a capitation system. Variants of both structures were subsequently adopted over the traditional salary model. In addition, the Government introduced increased flexibility into the budgeting process for institutions.

The Medicaid program in the United States also provides a variety of experiences in efforts to change provider behavior through financial incentives. The common goal of these efforts was to reduce overall program costs by adopting

[44] See Wlodarczyk (1987) for details.

[45] *Ibid.*

managed care programs.[46] Under a managed care system, patients have limited choice of providers, who are assigned to serve as gatekeepers and to provide general oversight to the care of particular patients. In most cases, the approach was to impose some form of financial risk sharing on the physicians or to reward them for providing cost effective care. Previously, physicians were reimbursed on a fee-for-service basis, under which there were no incentives for cost containment.

During the 1980s, demonstration projects in selected counties tested a variety of incentive structures. In Monterey, California, for instance, the Monterey Health Initiative opened in June 1983. Its physicians were paid on a fee-for-service basis, but with additional fees for case management. This approach did not involve the physicians in any financial risk sharing, and relied instead on what proved to be an inadequate management information system. The effort closed in bankruptcy in 1985.

Other demonstration programs were more successful. In Wayne County, Michigan, for instance, indirect financial incentives for physicians contributed to significant improvements in cost effectiveness. Any physician whose aggregate costs exceeded the 95th percentile had his or her case management fee held in escrow pending a review. If costs were deemed unjustified and the problem persisted, the physician would be excluded from future participation in the Medicaid program.

Demonstration programs in Santa Barbara, California, and in Jefferson County, Kentucky, instituted capitation schemes. If physicians' actual costs were below the capitation levels, the physicians were allowed to keep the profits; if costs exceeded capitation, the physicians were required to forego future payments to make up the loss. Similar structures applied to costs incurred for referrals to specialists. In Santa Barbara, hospital use rates during the program's first year fell 30 percent, while primary care physician use increased. In Jefferson County, hospital utilization rates had fallen by one-third and physician utilization increased by 16 percent during the year in which the program operated.

4.4 RELEVANCE OF THE DEVELOPED COUNTRY EXPERIENCE TO DEVELOPING COUNTRIES

Certain aspects of the developed country experience may be relevant to understanding the potential for provider incentives in developing country health systems. In particular, we hypothesize that individual providers and consumers of health care react to economic incentives similarly in developing countries as they do in developed countries.

Industrialized and developing countries exhibit many important differences, not only in the level of wealth, but also in their socioeconomic and cultural dimensions. Thus, we can only assume that the effects of changes in incentive structures on the efficiency, equity, and quality of care will tend to be similar in developed and developing countries. Our review of the literature and experiences in the developing world, presented in Section 5, tends to confirm

[46] Freund (1987) reviews this experience.

this assumption. This permits us to apply the results of incentive studies in developed countries, or at least to form our expectations. Since the developed country setting has been the subject of much more rigorous and detailed study in this regard, it may help to anticipate the outcomes of incentive initiatives not yet undertaken in developing countries.

The focus on physician behavior in developed country literature is also relevant for the developing country context, where physicians play the same clinical role but tend to play greater roles in administration as well. Usually, they are on the staff and payroll of the hospital, as opposed to being independent businesses which have admitting rights.

The developed country hospital literature, on the other hand, is somewhat less relevant to developing countries where for-profit hospitals account for a rather small proportion of the total supply of hospital services. Nevertheless, as developing countries move to consider alternatives to government-owned hospitals, the developing country literature on incentives and hospital behavior becomes more relevant.

Similarly, most developing countries lack complicated public health care finance systems, such as Medicaid and Medicare, the cost reimbursement aspects of which play a large role in developed country literature. Thus, developed country literature on physician incentives is more transferrable to the developing country context than is the literature on hospital incentives. However, as developing countries begin to consider alternatives to fully subsidized government production of hospital care, the experience in third-party hospital reimbursement drawn from industrialized nations may become more relevant.

The HMO literature in developed countries is increasingly relevant, as managed prepaid health care programs expand their coverage in developing countries. This currently applies more to Latin America and Asia than Africa. In fact, such programs spread widely through Latin America and the Caribbean during the 1980s. An AID-sponsored review of managed prepaid health care in Latin America[47] found a wide range of organizational structures and financing mechanisms. Most of the programs were associated with employer-provided health services.

In general, the incentive-related issues confronting developing country health systems are much more fundamental than the problems facing developed countries. Examples of disincentives abound. In general, physician and staff salaries in government health care facilities are well below market wages for equivalent positions in the private sector. This leads physicians and other health personnel to minimize the time spent at non-remunerative public facilities. Of course, the low salary is not the only incentive to "moonlight;" the lack of accountability also plays an important role. Thus, as already noted, even if government salaries are comparable to those of the private sector, government employees may still engage actively in private practice at the expense of their regular duties in government facilities, if doing so does not threaten

[47] Group Health Association of America, Inc. (1985) and Solari (1984).

their government jobs. Also, the managers of government facilities are often given little latitude to control financial and personnel issues. This simultaneously robs many of the incentive to improve the quality and efficiency of their facilities, and undermines their ability to do so if they were so inclined. In many places, revenues generated by facilities disappear into central treasuries, thus leaving the facilities with no apparent benefit from better performance. This litany of ills is sufficiently widespread and well-known so as not to require further elaboration.

In most developing countries, the physicians and other provider staff are civil servants. This generally means three things:

1) Their wages are generally low;
2) Rigidities in the compensation system for civil servants makes it difficult to implement monetary incentives; and
3) Assured tenure in public facilities implies that they are unlikely to be disciplined or dismissed in case of poor performance.

As a result of these constraints, the matter of motivating providers to devote full work days to the public facilities is problematic. Despite the fact that it is illegal in many countries for physicians to practice privately, the problem of moonlighting during "company" hours to increase their income is pervasive. Innovative incentive programs may help to ameliorate this problem in settings where salaries are fixed by civil service standards.

In this sense, incentives present themselves as a second-best tool to improve the situation at the margin. The global problem is one of lack of disincentives to poor performance or incentives for good performance. Innovative incentive programs present an opportunity to motivate greater effective labor, given the broader constraint of inadequate funding for salaries and equipment.

Other important differences arise from simply being in a low-income country. Small public health budgets often result in poorly stocked and poorly maintained health facilities. This problem has a dual effect. For providers, the conditions contribute to low staff morale, thus reinforcing the financial incentive to moonlight. For consumers, these conditions often make public facilities less attractive than alternative providers (that is, perceived quality is lower in the public facilities), thus lowering demand and redirecting to the private sector the revenues needed to improve the public system sector.

With these issues and distinctions in mind, we can turn to a review of provider incentive schemes that have been attempted in developing countries.

5.0 REVIEW OF DEVELOPING COUNTRY EXPERIENCE WITH PROVIDER INCENTIVE SCHEMES

The final piece of the foundation for our preliminary field work design is a review of developing country experiences with health care provider incentive programs. Unfortunately, actual developing country experience with these programs is limited, and few among the available examples are documented. Even more rare are examples for which data exist to permit before/after evaluations of the effect of provider incentives. In addition, all of the available examples from developing countries are from the public sector. Yet one of the lines of inquiry that is important for our research design is to examine the applicability of private sector incentive schemes to the public sector in developing countries.[48]

This chapter provides an overview of available examples. Many of the examples are anecdotal, based on verbal descriptions, their results unknown. The presentation is divided between monetary and non-monetary examples.

5.1 REVIEW OF MONETARY PROVIDER INCENTIVE PROGRAMS

PROSALUD, a USAID-funded non-profit organization in Bolivia, in conjunction with Management Sciences for Health, instituted an elaborate employee incentive program.[49] PROSALUD's ability to initiate such programs, despite the usual rigidities pertaining to public sector health worker salaries, was the result of an innovative management structure: the Government of Bolivia granted PROSALUD the role of managing and setting performance standards for public sector employees. In effect, the government subcontracted the management of its own employees to a non-profit, non-governmental organization.

PROSALUD is by far the best documented example available, and is recounted here in some detail. In fact, PROSALUD implemented three distinct incentive schemes. The overall objectives of these incentive programs were:

1) To encourage increased use of the facilities by linking staff incomes to health center income; and
2) To spread the facility's financial risks among the employees. Each strategy was thus intimately bound up with PROSALUD's goal of financial self-sufficiency through cost recovery.

The first program was a model for medical specialists, including pediatricians, gynecologists, and dentists. The PROSALUD health centers were individually unable to support full-time specialists. PROSALUD thus instituted an arrangement through which specialists were contracted on a part-time basis in which they and PROSALUD shared the income risk. The specialists received no salary from PROSALUD. Instead, they were allowed to keep 50 percent of whatever revenues they generated from patient billings, while PROSALUD received the other

[48] Of course, there are non-profit and commercial private providers in developing countries which use incentive systems. To the best of our knowledge, these have neither been documented nor evaluated.

[49] The discussion of PROSALUD is drawn from Management Sciences for Health (1989a) and (1989b).

50 percent. This arrangement has proven financially beneficial for both the health centers and the specialists.

PROSALUD also initiated an incentive program for physicians at new health centers. In order to reduce the start-up costs of the new facilities and to motivate physicians to draw in more patients, PROSALUD pays the physicians only 80 percent of the average salary for their general practitioners. Additional physician salaries are calculated as a percentage of the revenues generated from new patients. When this amount equals the average salary for general practitioners, the base salary of the physicians at the new health centers is increased to the average.

In terms of the payment structures outlined in Chapter 4, this could be defined as a combination of salary and fee-for-service. The incentive for these physicians is to attract as many patients as possible.[50] We predict that this incentive structure will have negative consequences for economic efficiency, since there is no incentive to consider the costs of inputs. Since the demand for health care is also a function of perceived quality, physicians seeking to attract new patients have the incentive to improve quality.

The third incentive program established by PROSALUD provides the richest example, and we present it in the greatest detail. This scheme created a new system for allocating annual salary increases for health center staff, replacing automatic increases with increases based on individual performance.

The existing raise policy had been for all staff to receive automatic yearly increases regardless of the financial status of the facility and the employees' performance. The new system made annual pay increases contingent on individual productivity, measured by the number of patients seen. Behind this contingency was a perception on the part of PROSALUD managers that there was under-utilization of both curative and preventive services. The level of bonuses approximated the previous levels of automatic increases. Consequently, under the new system, the average worker had to work harder than before to earn the same level of income.

PROSALUD management evaluated each employee along a scale from 0 to 100. To be eligible for a bonus, employees had to rate at least 70, which signified "satisfactory but needs improvement." In addition, the health center had to perform a specified minimum number of preventive services to qualify for the bonus. This last condition was introduced to promote delivery and consumption of under-utilized preventive services.

For each health center, PROSALUD calculated a base level income as the average income generated from initial visits during the past eight months. Thirty percent of the revenue generated in excess of the base level for a given health center was held in reserve to buffer slow months. Any money left in the

[50] But only up to the point where they reach the average salary. From then on, there is no further incentive to attract patients.

reserve at the end of the year was divided between the physicians, other health center employees, and PROSALUD.

The expected results from this incentive program were:

- Increased productivity among employees and thus a higher level of financial self-sufficiency for PROSALUD;
- An average increase in staff salaries of 10 to 15 percent; and
- The provision of a significant number of preventive services.

At first, health center employees were not in favor of the program, which meant the loss of guaranteed salary increases. After several months of operation, when bonus levels began to grow, the program gained in popularity. Yet, after one year of operation, bonus levels plateaued, and the program's popularity waned.

The program's effect on PROSALUD's financial self-sufficiency was quite positive. Prior to the program's initiation, the overall level of self-financing by health centers was 78 percent; six months into the program, the level had grown to 98 percent, and 16 months into the program, the level had grown to 109 percent. Total income generated in the first year increased by one-third.

With regard to anticipated increases in average salary levels, the results were somewhat disappointing. Physician salaries in the first year increased by an average of 10 percent, while staff salaries increased by only eight percent. These increases were in nominal levels (i.e., not adjusted by inflation) and were lower than the previous rate of automatic increase, which was 18 percent. In addition, virtually all of the salary increases were gained by employees at urban health centers. Rural health center workers received virtually no increase.

Concerning the mix between preventive and curative care, the effect of the incentive program was to increase the number of curative cases without significantly decreasing the number of preventive services. Thus, there was a relative shift in emphasis towards curative care, which may signal a potential efficiency loss for the system.

All of the remaining examples in this chapter are much less detailed than the case of PROSALUD. Indeed, for most of the remaining examples, no information was available regarding outcomes. These examples are also drawn from a diverse set of contexts, ranging from hospital care to insurance sales.

The issue of preventive versus curative care was explicitly addressed in employee incentive schemes in Benin and Guinea, implemented under the auspices of the Bamako Initiative.[51] Health center staff bonuses there were tied to the provision of preventive services. Health personnel received 200 FCFA (about US$ 0.75) for each child correctly and completely vaccinated within one year of birth and each pregnant woman receiving at least three prenatal consultations, two

[51] Knippenberg, et al. (1990).

tetanus toxoid vaccinations (appropriately spaced), and who delivered in the health center. Staff also received 50 FCFA for each curative consultation.[52] The purpose of this program was to motivate personnel to increase coverage of preventive services, while curative care incentives were intended to generate health center revenue. A similar program was implemented in Guinea. In both cases, incentives were paid only for those cases treated according to pre-defined norms to discourage over-prescription of services.

Several examples also present themselves from Zaire. In 1982, the Government of Zaire restructured its centralized health system into over 300 financially semi-autonomous health zones. Each health zone consisted of a central reference hospital and numerous health centers, and operated as a de-facto parastatal, non-profit enterprise.

Each zone was free to adopt its own financing mechanism. In general, the zones adopted one of two systems: either the health centers were permitted to keep all of their revenues (again, attained through cost recovery) but were required to pay a supervision fee to a zone administrative and technical office, or the central office collected all revenues and paid personnel at the health centers.[53] In both cases, the local community boards oversaw the finances and quality of the services provided.[54]

The precise disposition of funds collected through cost recovery is, in fact, a critical consideration for employee incentive programs. A pervasive theme among incentive programs linked to cost recovery is the notion that some or all of the funds collected should be retained at the facility level. When 100 percent of the funds collected "disappear" into some centralized treasury, the incentive value for facility staff is largely lost. Similarly, if government subsidies to health facilities are reduced by an amount equivalent to the revenue raised from cost recovery, the facility employees have little incentive to generate funds from cost recovery. In the Zaire health zone example, the fact that funds not retained at the health center at least remain in the health zone tends to mitigate this loss; how much so depends on the specific formula by which the health zone administration reallocates the funds.

A recent study of health insurance in Zaire provides another example of a staff incentive system.[55] In Zaire's Bwamanda zone, prepaid health insurance functioned well and benefitted from an enrollment rate of over 60 percent of the population in 1988 and 1989. In contrast, health insurance enrollment rates in Bokoro zone were less than 10 percent. This striking difference may in part be due to the existence of an incentive program in Bwamanda that was lacking in Bokoro. Health workers in Bwamanda received a commission of three percent of the premiums they collected, while workers in Bokoro, had no incentive to increase enrollment in their zone. Moreover, in Bokoro, profit-making health centers were

---

[52] Reported cash incentive levels are from Benin.

[53] The government continued to pay salaries to some of the health zone personnel. Steep inflation, however, meant that these salaries quickly became a nominal payment. Most of the income of the health zone employees then came from cost recovery revenue.

[54] Bitran, et al. (1987).

[55] Shepard, et al. (1990).

permitted to retain only 15 percent of their profit for investments and 10 percent for discretionary use by the local health committee. There were also no penalties for poor performance.

In Puerto Rico, the municipality of San Juan used provider incentives as part of its effort to improve the cost-effectiveness of AIDS care.[56] AIDS care, 86 percent of which was inpatient care, was threatening to swamp San Juan's health care system.

In 1988, the municipality contracted with a private, non-profit organization to provide comprehensive AIDS care. While the same level of government funding was devoted to the new system, the non-profit entities shifted their emphasis towards preventive activities and made an effort to reduce inpatient care. Doctors were paid three times the going salary at the Municipal Hospital, and nurses' salaries were increased 67 percent. Additional non-monetary incentives included support for education and opportunities for research. Indeed, tenure at the non-profit institution was contingent upon being responsive to patients' needs.[57]

In return for these incentives, physicians were assigned particular cases, and were required to be accessible 24 hours a day (whereas under the previous system, physicians did not follow particular cases and were available only during the day). As a result of these changes, the mean length of AIDS inpatient hospital stays fell by 47 percent and the average annual per capita cost of inpatient AIDS care fell by 74 percent. Moreover, the study concluded that the quality of AIDS care improved during the course of these reforms.

The remaining examples are either more fragmentary or consist of plans rather than experience. The Ivory Coast, for example, attempted to address the problem of motivating physicians to stay at their assigned public health facilities during clinic hours. Most of the physicians minimized their hours at the public facilities in order to pursue illegal private practices. In response to this problem, the MOH passed a law legalizing private medical practice and requiring physicians to sign a declaration that they would work exclusively in either the public or the private sector. Those remaining in the public sector would receive salary increases, and those choosing private practice would be on their own. Virtually all of the physicians signed the pledge to stay in the public sector, received their salary increases, yet continued to maintain private practices just as before. The government was unable to enforce the pledges and, therefore, the policy was ineffective.[58]

In other instances, health center officials have used revenues from user fees to provide unofficial pay supplements to staff and hire additional staff. This was the case in a number of health centers in Togo, where ad hoc community

[56] Kouri, et al. (1991).

[57] Authors' interview with Donald Shepard (HIID), September 3, 1991.

[58] Authors' interview with Mead Over (World Bank), August 20, 1991.

organizations and health center staff also used cost recovery revenues to purchase medical supplies.[59]

In Chile, it was found that physicians were performing what appeared to be more cesarian deliveries than warranted clinically. In part, this may have been in response to a demand by women for the procedure. This illustrates the problem described above relating to suppliers' versus demanders' perceptions of quality. The physicians were paid on a fee-for-service basis. Since their salaries increased with the number of services performed, they had the incentive to compete for patients. If there was a demand for cesarian deliveries, physicians might have been tempted financially to meet that demand.

In response to that problem, an incentive program was initiated through which physicians would be given more equipment and resources if, as a group (at a given hospital) they met certain resource allocation targets (i.e., performed fewer cesarian deliveries). The program failed because the physicians individually gained more by continuing to perform the cesarians. Their perceived individual benefits under the group incentive were less attractive.

5.2 REVIEW OF NON-MONETARY PROVIDER INCENTIVE PROGRAMS

There are relatively fewer examples in the category of non-monetary incentives for health providers. One notion has to do with the benefits of creating an institutional environment in which managers have the ability and the incentive to motivate staff to be more productive. In particular, it has been observed in several countries that the constraints of civil service regulation prevent health facility managers from exercising authority over personnel and financial matters. Drawing on experience in Kenya, Indonesia, and Egypt, Stevens has observed that this constraint prevents managers from improving the performance of their facilities.[60]

The incentive Stevens proposes is to give these managers greater control over hiring and firing decisions and greater latitude in disposing of revenues and resources at their facilities. While these changes might lead to the creation of monetary incentive schemes for staff at those facilities, these non-monetary incentives to managers (e.g., increased authority, prestige, and ability to motivate staff) are seen as prerequisites.

Implementation of the Bamako Initiative in Guinea and Benin illustrates the positive incentive effects of increased supervision of health facilities. Regular visits by provincial health directors to health centers was found to contribute to improved quality of care and staff behavior.[61] The supervisors observed curative consultations, with the intention of determining whether care was being provided according to established norms. A review of this experience found that this supervision sensitized staff to the negative quality implications of over-prescription, despite the opportunity it affords to increase revenues.

---

[59] Authors' interview with Marty Makinen (Abt Associates), September 5, 1991.

[60] See Stevens (1984), (1986), (1989), (1990), (1991) on Kenya, Indonesia, and Egypt.

[61] Knippenberg (1990).

Non-monetary incentives of another type have been attempted in response to the problem cited earlier of motivating physicians to spend appropriate amounts of time at their assigned public facilities. Rather than making private practice (i.e., moonlighting) illegal, several countries have initiated policies that permit physicians to see private patients at the public facilities during evening hours. The goal of this approach is to keep the physicians at the facilities. In some cases, the physicians are required to pay nominal rent to the facility for this privilege. Initiatives of this type are under way in several sub-Saharan African countries, including Ghana, Tanzania, Zambia, and Sudan.[62] In order to control the whereabouts of its physicians, Kenya's Kenyatta National Hospital (KNH) conducted a feasibility study of renovating a wing where its physicians could treat their private patients. The staff physicians using the private wing would pay rent to KNH as well as a fee for the use of KNH's services.[63] The greatest pitfall in these programs, however, is preventing physicians from diverting patients and resources from their public into their private practices.

Another example of non-monetary benefits comes from Chile. In that country, physicians with specialty training receive significantly greater fees than general practitioners. As a result, specialty training is much sought after by physicians, though there are places for less than 10 percent of those who seek them. The Government of Chile has introduced a program in which it guarantees specialty training to physicians who agree to serve for a five-year period in medically under-served rural areas. This program is said to work quite well.[64]

[62] Author's interview with David Parker (UNICEF), October 18, 1991.

[63] Hildebrand, S., "Kenya Trip Report," HFS Project, Bethesda, MD, 1990.

[64] Author's interview with Karen Lashman (World Bank), August 28, 1991.

6.0 SUMMARY AND CONCLUSIONS: PRELIMINARY CONSIDERATIONS FOR FIELD WORK

This paper has presented a conceptual discussion of provider behavior, work incentives, and performance of government health systems. The analysis was used to support the hypothesis that provider incentives can be used to improve productive efficiency in government health services. Incentives developed to promote greater efficiency in health services production, however, can also affect the quality of care and the equity of the system. Nevertheless, it was shown that it is possible, at least in theory, to develop incentive packages which will mitigate—or even reverse—the negative effects on equity and quality of efficiency-seeking incentives.

A review of the literature on provider incentives and health system efficiency was conducted using sources from both developed and developing countries. The review from the industrialized world finds that the literature focuses mainly on economic incentives to physicians and hospitals, the latter primarily through various reimbursement mechanisms. The relevance of the industrialized country literature to the developing country context is established. With regard to the experience of developing countries with provider incentives, we find a small number of documented sources. However, interviews with several experts reveal great interest in this line of research.

A primary purpose of this paper was to set the conceptual foundations for the design of one or more field work activities to be undertaken by HFS in the area of provider incentives and productive efficiency. The following section is a discussion of methodological issues associated with the measurement of efficiency in health services production and with the effects of incentives on efficiency, quality, and equity. The final section presents the preliminary design of three phase-two research activities which will explore the issue of incentives and provider efficiency.

6.1 METHODOLOGICAL ISSUES PERTAINING TO THE MEASUREMENT OF EFFICIENCY

The measurement of efficiency of health services poses various methodological problems. Fortunately, the relevant literature presents several techniques for overcoming these empirical obstacles. As with most research initiatives, greater accuracy in the measurement of efficiency calls for more sophisticated—and thus costlier—measurement techniques. The HFS Project is currently completing a review paper which explores in detail the issue of efficiency measurement in health systems.[65] Like this document, that paper also constitutes the first phase of a larger, major applied research initiative whose purposes are to advance knowledge about measurement techniques and provide information about the levels of efficiency of selected health care providers in developing countries. A study of provider incentives and production efficiency necessarily involves the measurement of efficiency. For that reason, and in order to reduce overall research costs for HFS, it is proposed that as much as possible of the field work on efficiency measurement be conducted in conjunction

[65] Bitran, R., "Technical and Economic Efficiency, Phase 1: Review of Concepts and Literature, and Preliminary Field Work Design," HFS Project, 1992.

with the empirical work on provider incentives and efficiency. The remainder of this section is a brief discussion on efficiency measurement issues that draws from the previously referenced paper. Discussions about the feasibility of empirically identifying links between provider incentives and efficiency and about the measurement of equity are also provided.

6.1.1 Methodological Problems in the Measurement of Efficiency

The measurement of efficiency in health care services presents three important methodological problems:

- Meaningful comparisons of efficiency among providers must be made while controlling for quality of care. Otherwise, lower costs attributable to inferior quality could erroneously be ascribed to higher efficiency, and vice-versa.

- Variations in case mix (i.e., the types and complexity of the medical cases treated) must also be controlled for when comparing provider efficiency. Failure to do so may yield the misleading result that those providers treating more complex cases appear less efficient than those treating simpler cases.

- Price variations and distortions make comparisons of costs among providers difficult, particularly between private and public providers.

6.1.2 Techniques for Measuring Efficiency

There are several methods for measuring efficiency, but all of them require control for quality, case mix, and price distortions. The measurement techniques available include:

1) Accounting-based studies of efficiency;
2) Deterministic studies of efficiency that use Data Envelopment Analysis (DEA) or related linear programming techniques; and
3) Econometric studies of efficiency which estimate health provider cost or production functions and derive associated indicators of efficiency.

6.1.3 Controlling for Quality Variations

Quality differences can be controlled in several ways and with different levels of rigor. Possible methods include:

1) The measurement of quality by contrasting the technical performance of the provider with norms of treatment;
2) Quality measurements that are based on patient satisfaction information obtained through surveys;
3) Quality assessment that is based on various measures of health outcomes using information on patients' views about seriousness of discomfort and self-reporting of post-treatment problems, and medical charts which provide general information on complications, general health status, symptoms, and complaints; and

4) Combinations of any two or all of the above methods.

6.1.4 Controlling for Case Mix Differences

Case mix differences can also be controlled in various ways. Available methods in the literature include:

1) At a micro level, the choice of a sample of health procedures that are relatively simple and homogeneous (i.e., with little possible variability in complexity), such as hernia repair and tubal ligation;
2) At a more aggregate level, the choice of health facilities or providers which treat a relatively similar set of medical problems; and
3) The use of econometric techniques which statistically control for variations in complexity and their effect on cost and, thus, efficiency.

6.1.5 Correcting for Input Price Distortions

Input price variations among providers obscure the interpretation of comparative analysis of efficiency. In particular, when factor price variations exist, no definite inferences can be made about economic efficiency from a study of technical efficiency. In addition, differences between factor transaction prices and the actual social cost of those resources also limit the validity of studies of economic efficiency. The use of social costs for the price of production resources is a solution.

6.1.6 Isolating the Effects of Incentives on Efficiency

The U.S. literature on production efficiency has generally dealt with this problem in two steps. The first step consists of measuring provider efficiency with one or more of the several methods available. In the second step, efficiency indices are regressed against a series of independent variables which are presumed to influence efficiency. These variables have typically included dummy variables associated with the form of ownership of the facility. This method requires large enough samples of facilities to permit statistical estimation.

An alternative to a statistical study of incentive effects on efficiency is a descriptive study. This type of research can describe existing incentives and measure provider performance, but it cannot statistically establish causal links between incentives and performance. This, however, should not preclude HFS from pursuing this type of initiative. In fact, there is so little research in this area in the developing world that additional contributions, even if they do not reach the stage of statistical certainty, would be welcome.

6.1.7 The Measurement of Equity

Equity in service provision can be measured in at least two ways:

1) By conducting patient surveys, collecting information about payments, and contrasting patient characteristics with characteristics of the overall population to assess whether particular population groups are less likely to obtain care; and

2) By conducting household-based studies of health care demand and utilization and assessing any differences in both measures among different population groups.

6.2 PROSPECTIVE FIELD WORK ACTIVITIES AND PRELIMINARY RESEARCH DESIGN

Three opportunities for phase-two field work are envisioned: a study of personnel incentives and performance in ambulatory government health facilities in Niger; a study of promoter incentives and performance in private voluntary organizations (PVOs) in the Dominican Republic; and a comparative study of public and private provider incentive systems in government hospitals in Egypt. This section briefly provides background information on these three activities and proposes a preliminary research design.

6.2.1 Study of Personnel Incentives and Performance in Government Ambulatory Health Facilities in Niger

The Niger Health Sector Support Grant, a Non-Project Assistance Grant funded by AID, has completed the design of a major field test experiment for cost recovery for ambulatory care in that country. Test activities are expected to begin in June 1992. Abt Associates Inc. has been involved in grant activities since its inception, as a grant subcontractor. In that capacity, Abt has lead the research design effort for the cost recovery pilot tests. These were proposed by senior government officials to assess the pros and cons of various cost recovery options before considering cost recovery for national implementation.

The tests offer a unique opportunity to conduct major applied research in the field for several of HFS's major applied research topics, including provider incentives and efficiency. In effect, the tests will deal with issues of means testing, social financing, quality of care, efficiency in consumption, and efficiency in production, in addition to provider incentives.

The tests will involve the MOH ambulatory facilities of three districts of the country. Health care in MOH facilities is currently free of charge. One district will adopt a fee-per-episode cost recovery system; another will adopt a head tax combined with a small copayment per illness episode; the third district will be a control setting. Cost recovery will be implemented for curative care only and is expected to cover the full cost of pharmaceutical products and medical supplies. The Government of Niger is expected to maintain its current levels of subsidization to the test facilities, including the payment of personnel salaries.

Household surveys will be conducted before the introduction of cost recovery and a year later, to assess any demand changes. The financial and technical performance of MOH health facilities will also be assessed through monthly facility surveys.

Exhibit 6-1 presents a workplan for the overall pilot test initiative and Exhibit 6-2 presents the associated level of effort. The tests and their evaluation are expected to take place over a period of about two years and to use about 5,000 person-days, or 227 person-months. It is assumed in the exhibits

that HFS would provide the technical assistance staff required to run the tests and the staff to evaluate and research various aspects of the tests. It is also assumed that HFS would use the tests to conduct major applied research in at least four major areas, including, possibly, provider incentives and efficiency, means testing, social financing, and quality of care. The level of effort described, therefore, would be shared by at least four major field research activities.

While the design of the tests has recently been completed, several important issues remain to be resolved, particularly with regard to provider incentives. The research design proposes that health facility personnel be provided with monetary incentives calculated as a fixed percentage of either the total revenue of the facility or the difference between total revenue and the total cost of pharmaceutical products. It is hoped that the incentives would supplement personnel salaries by 10 to 30 percent. Also, incentives would be conditional on attaining a minimum level of delivery of preventive services, as was done by PROSALUD in Bolivia.

Since health care and drugs are currently provided free of charge in government facilities, it is not known how cost recovery will affect demand. Prices of care will be higher, but so will be health care quality given the improved availability of drugs and medical supplies that will be possible with cost recovery. The facilities' revenue, therefore, cannot be anticipated, let alone the magnitude of the bonuses that could be paid to the personnel.

6.2.1.1 Study Goal

The goal of the study is to determine how monetary incentives to health facility staff affect the facility's productive efficiency, equity in delivery, and quality of care.

6.2.1.2 Methodology and Workplan

Exhibits 6-1 and 6-2 describe the workplan and level of effort for the entire pilot test undertaking. The study of provider incentives and performance will take place as part of the tests. The activities and level of effort are thus imbedded in the charts. A preliminary methodology for studying incentives and performance is sketched next, even though considerable uncertainty remains about whether the incentive system proposed is at all acceptable to the Government of Niger.

About five months will elapse from the time the tests begin to the introduction of cost recovery. Decisions about whether monetary or other types of personnel incentives will be adopted, and about the types of systems to be adopted, if any, will be made during this period. Facility questionnaires will be designed to collect from participating health facilities information on output levels, input prices, availability of drugs, revenues, provision of free care, incentives paid, and production costs. Enumerators will be trained to collect facility data on a monthly basis.

Exhibit 6-1
Niger Cost Recovery Pilot Tests - Gantt Chart

Pilot Test Implementation Plan - 1992

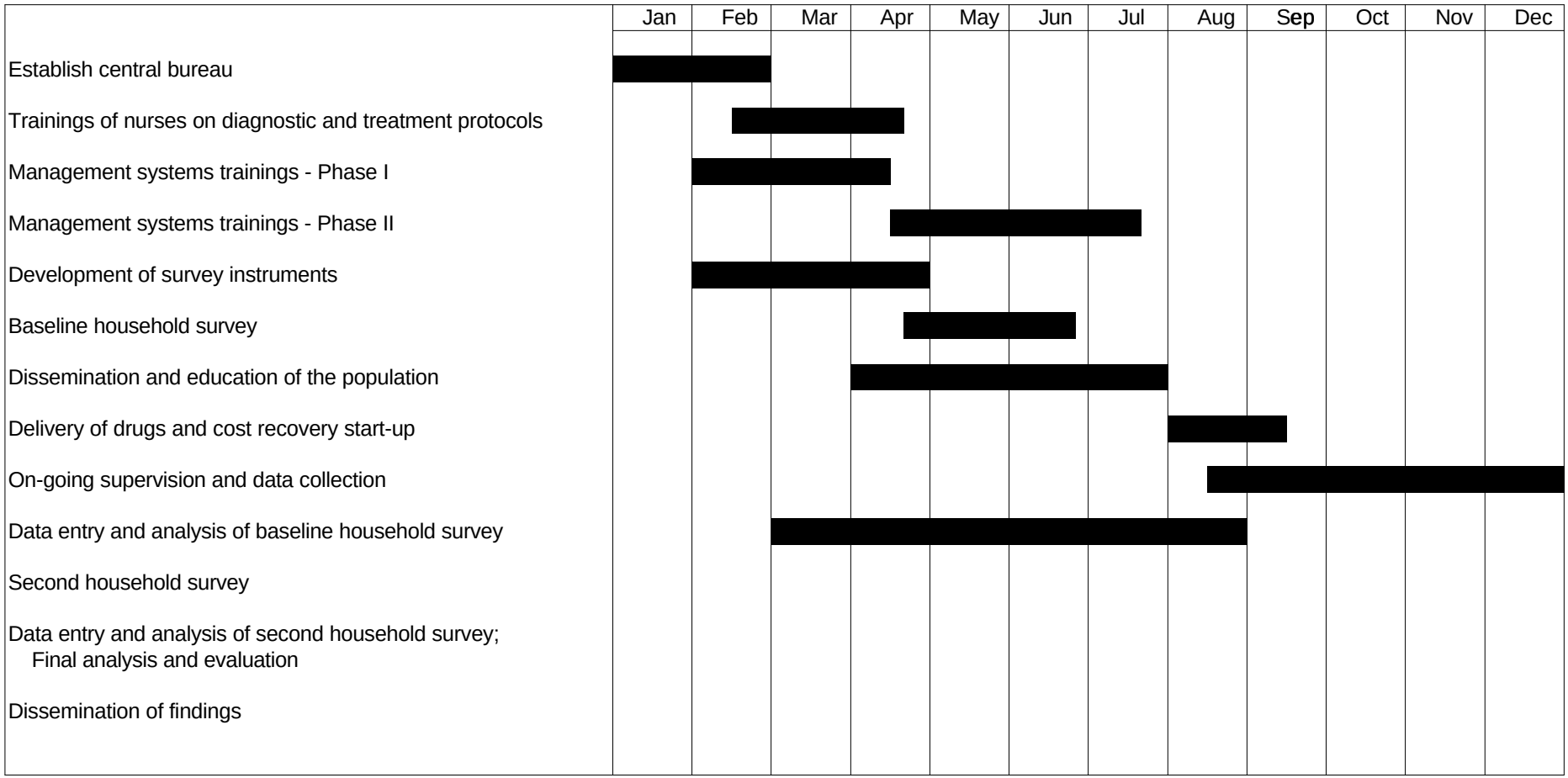

Exhibit 6-1
Niger Cost Recovery Pilot Tests - Gantt Chart
(continued)

Pilot Test Implementation Plan - 1993

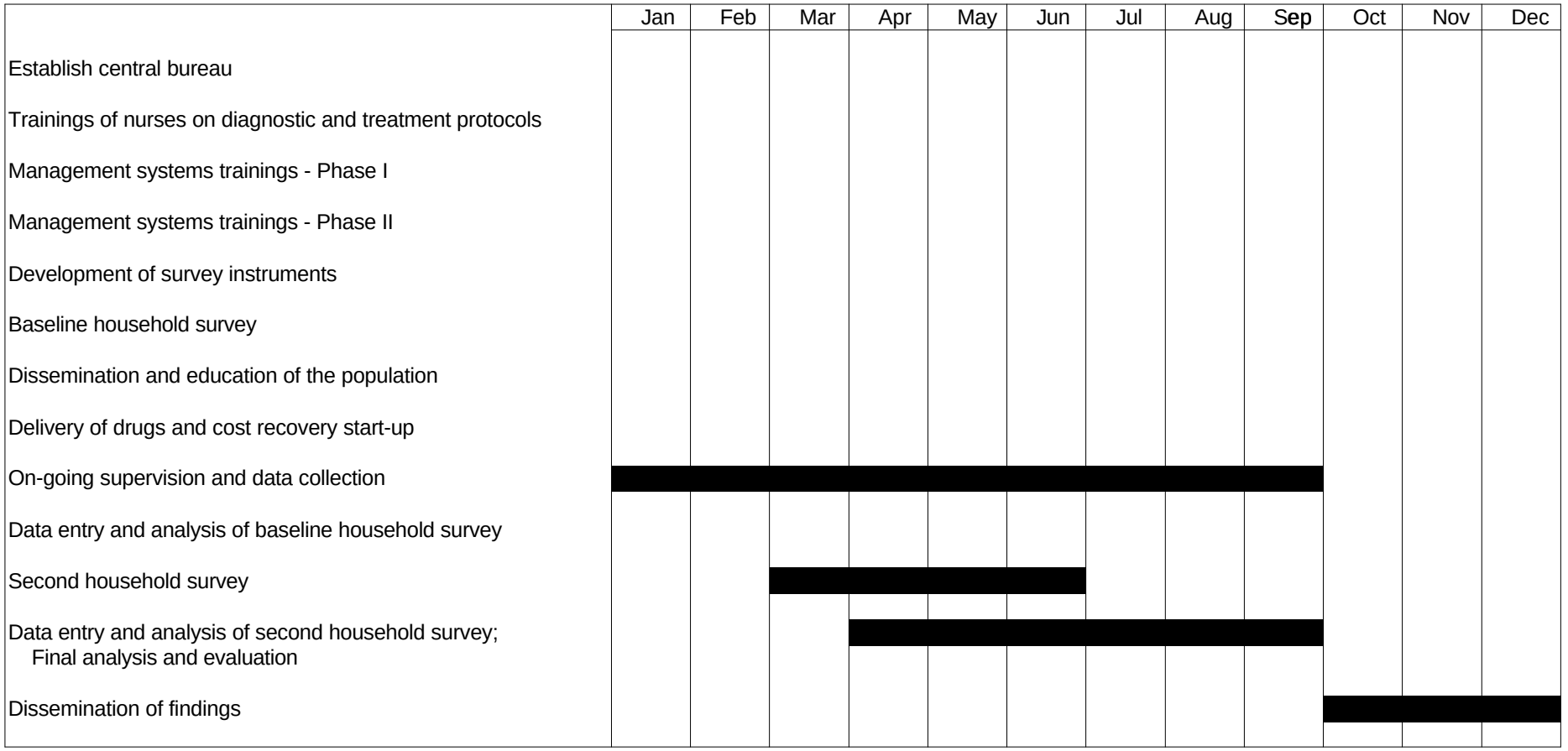

Exhibit 6-2
Niger Cost Recovery Pilot Tests - Estimated Level of Effort

Niger Cost Recovery Study Days of Effort by Task	Loaded Daily Rate	Total Days (Effort)	Total Cost	Set-up Central Bureau	Nurses Training	Mgmt. Systems Training I	Mgmt. Systems Training II	Survey Instru-ments	Develop Baseline HH Survey	Popul. Educ.	Dist. Drugs/ CR Start-up	Super-vise/ Collect Data	Data Entry/ Analysis	2nd HH Survey	Analysis of 2nd Survey/ Final Anal. & Eval.	Dissemi-nation of Findings
HFS Team Director	832.00	12	9,984												10	2
HFS Director of Applied Research	832.00	64	53,248	5		2	2	5							40	10
HFS Task Manager	450.00	330	148,500	22		22		22								22
HFS Analyst	450.00	160	72,000			20		30							100	10
HFS Chief Trainer	200.00	69	13,800			44		25								
HFS Data Entry Chief	200.00	100	20,000										50		50	
HFS Chief of Data Collection	200.00	40	8,000						40							
HFS Subtotal		775	325,532													
Enumerators	30.00	2760	82,800					240	960			360		240		
Trainer in Management Systems	30.00	198	5,940			88	110									
Health Facility Supervisors	30.00	360	10,800									360				
Data Entry Staff	30.00	500	15,000										250		250	
Drivers	30.00	300	9,000					30	240					30		
Secretary	30.00	0	0													
Translators	150.00	60	9,000					20			40					
Local Subtotal		4178	132,540													
TOTAL:		4953	458,072	27	0	176	112	372	1240	0	40	720	300	270	450	44

Health facility efficiency will be assessed using either ratio analysis or econometric estimation of cost or production functions and associated measures of efficiency. The choice of the method will depend on resource availability. Quality of care can be assessed in two ways: through patient surveys at the facility or through technical reviews of provider compliance with norms of care. Equity in the provision of care can also be gauged, either by analyzing health care demand or, as a second best, by studying how facilities administer free care. The effect of incentives on efficiency can be studied by regressing various efficiency measures against several variables believed to affect efficiency, including the type and size of the incentive. Simpler but more ambiguous methods for studying incentive effects on efficiency consist of comparing efficiency indicators with incentives paid.

6.2.2 Incentive Systems and Rural Health Promoters in the Dominican Republic: The Effect on Recruitment, Coverage, and Quality[66]

Since 1989, the Child Survival Project (PSI) has supported the provision of mother-infant health services (MCH) through a number of health sector PVOs. The project focuses on providing a limited number of child survival interventions in three regions. Two regions are located near the Haitian border: Barahona (Sanitary Region 4) and San Juan (Sanitary Region 6). PSI has provided subgrants to 10 PVOs operating in these regions.

All PVOs depend on semi-volunteer health workers to provide services at the community level. Known by a variety of labels (e.g., promoters, facilitators, health assistants, and extension agents), these workers are generally women recruited from the communities that the PVOs have targeted. In short, the promoters are the PVO's front-line health workers.

University Research Corporation (URC)/PSI focuses on three health education interventions performed by promoters: diarrhea management, birth spacing, and breast feeding and weaning practices. Some of the PVOs participating in the project, however, require that the promoters execute additional functions such as child growth monitoring and control of infectious diseases.

On paper, PVOs refer to the promoters as volunteers. In practice, depending on their philosophical and religious orientations, most PVOs use one or more incentive mechanisms to motivate their promoters. Examples include: cash transfers, in-kind gratuities, commissions, full-time salaries, and spiritual guidance.

Promoter incentives have become a controversial issue among PVOs and a source of tension between PVOs and their promoters. A number of PVOs profess that training, prestige (within the community), and spiritual fulfillment should provide sufficient motivation for the promoters. This group strongly rejects the provision of cash incentives.[67] According to these sources, voluntary service to the community, a theme that is core to their development approach, has eroded

[66] This research proposal has been written by HFS staff member Gerard La Forgia.

[67] A number of these PVOs are affiliated with religious institutions.

in recent years as promoters seek to sell their services to the highest bidder. They refer to this phenomenon as the "professionalization" of community promoters.

Others hold the view that incentives have been part and parcel of promoter-based delivery systems in the Dominican Republic since the early 1980s. At that time, the MOH launched a rural health program in which community health promoters were (and still are) paid token cash incentives. Moreover, although it is unknown whether promoters as a group are better off than the clients they serve, PVO field staff state that they are generally poor and seek to improve their life situation. Consequently, promoters tend to use their training and skills acquired on the job to secure additional income. In some areas, promoters have accumulated several years of experience participating in PVO and government programs health (and non-health) projects. By this account, paying for their services is the best way to keep them active in community development. In short, several PVOs are convinced that cash incentives are here to stay and are a requirement to recruit and retain quality promoters.[68]

Expectedly, promoters pressure PVOs for more cash and in-kind incentives. The promoters often do not share the voluntary conviction set forth by their central-office superiors. They point out that fulfillment of their task and coverage responsibilities can require up to 10 hours per week. To be sure, at a meeting of regional PVO staff in May, 1990, it was agreed that 10 hours per week was a benchmark for promoter time requirements regarding MCH tasks. An author of a recent PVO evaluation determined that under current MCH task guidelines, a promoter would need to dedicate an average of 17 hours per week to cover families within her catchment area.

Absent from this debate is any discussion of the effect (if any) of incentive systems on coverage, quality, and continuity of promoter-based services. Another issue is the relationship between incentives and promoter recruitment and turnover. Family-per-promoter and promoter-per-supervisor ratios can also affect promoter performance. This suggests another issue: the cost trade-off between (higher) incentives and (lower) family-promoter and promoter-supervisory ratios (e.g., recruitment, training, supervision costs). Exhibit 6-3 compares incentive mechanisms and estimates family-per-promoter and promoter-per-supervisor ratios for seven PSI-funded PVOs providing MCH services in rural areas of Barahona and San Juan. The exhibit demonstrates considerable variation in incentive systems and family-promoter and promoter-supervisor ratios.

[68] Another incentive not reported by the PVOs is worth mentioning. All PVOs hire supervisors to oversee promoter activities. Most supervisors are drawn from the promoter ranks. These full-time, salaried positions are highly coveted and may represent an important incentive to the promoters. The opportunity for employment may have a positive effect on performance levels and promoter retention.

6.2.2.1 Hypotheses

The purpose of the research is to determine the effect of incentive systems on promoter performance. Performance can include: real population coverage, continuity of coverage, and quality of care provided. Incentives also may affect the ability of a PVO to recruit and retain experienced promoters, which in turn may influence knowledge levels, and ultimately, performance. Confounding variables can include: family-per-promoter ratios, promoter-per-supervisor ratios, number of different tasks assigned to the promoter, estimated number of hours required to complete these tasks, size of catchment area (geographical dispersion), family size, and the total number of people receiving certain services (e.g., malnourished children requiring monitoring). As discussed later, the study will attempt to control for a number of these variables through the research design.

Exhibit 6-3
PVO Child Survival Programs:
Promoter Coverage, Supervisory, and Incentive Systems

PVO	AVG. NO. FAMILIES P/PROMOTER[a]	AVG. NO. PROMOTERS P/SUPERVISOR[a]	FORMAL INCENTIVE POLICY	MONTHLY CASH INCENTIVE (in RD $)	TYPE OF IN KIND INCENTIVES
MUDE	60	18	yes	$100	NA
FH	43	21	no	0	medical care, food, clothes, etc. on "as needed" basis
AED	55	12	no	0	school supplies, medicine, and household items on "as needed" basis
CARITAS	43	10	yes	50[b]	recognition awards and gifts
SSID	21	18	yes	50	NA
FUDECO	40	20	yes	60	
CARE	6 clinics; 865 benef.[b]	18	yes	3,315[d]	motorcycle; health insurance

[a] Based on PVO estimates of coverage. Real coverage is probably lower.
[b] Promoters also given access to credit.
[c] CARE "extensionists" cover target groups (e.g., infants and pregnant women) within the catchment area of SESPAS rural clinics.
[d] Wages, per diem, and gas allowance.

Source: La Forgia and Heinig, 1992

Incentives will positively affect promoter performance in three ways:

- Promoters who receive greater incentives demonstrate higher levels of coverage and quality of service.
- Promoters who receive greater incentives had more years of experience prior to recruitment. Promoters with more years of experience demonstrate superior knowledge levels (interaction effect).

- Promoters who receive greater incentives and have more years of experience demonstrate superior levels of coverage and quality (interaction effect).

6.2.2.2 Methodology

The study will be conducted in Barahona (region 4) and San Juan (region 6). The regions are coterminous and relatively equivalent in terms of several relevant variables (CENISME, 1989, 1990). Exhibit 6-4 compares several summary characteristics. Ten PVOs in these regions provide services to approximately 300 rural and urban communities. Generally, each community is served by one promoter. More dispersed areas are served by two or more.

Exhibit 6-4
Summary Characteristics of Two Sanitary Regions

CHARACTERISTIC	YEAR	REGION 4	REGION 6
KM^2	--	6732	7779
(INHAB./KM^2	1981	40.3	57.6
AVG. NO. OF INHAB./RURAL COMMUNITY	1981	258	274
INFANT MORTALITY	1990	82	87
% CHILDREN MALNOURISHED	1990	39	45
% VACCINATION COVERAGE (MEASLES AND POLIO)	1989	62	61

Source: IEPD (1982); CENISME (1989, 1990, 1991).

The methodology will incorporate an "ex post facto" design. In this quasi-experimental design, each group of promoters that has experienced a particular incentive system, X_1, is compared to other groups that have experienced other incentive systems, X^i, to measure the effect of X on performance. The key to this design is the incorporation across all groups of relevant variables that also may affect performance, such that the groups are equivalent if it had not been for X.[69] In other words, the units of analysis across all groups, promoters/communities, will be stratified into homogeneous blocks in which they are matched on attributes that can contribute to a performance effect. To account for the effects of self-selection of promoters into a treatment group (PVO incentive systems), the design will measure the treatment interaction

[69] In addition to the technical problems inherent in this design (later described), the PVOs and PSI may prefer that we do not attempt to isolate the effects of the incentive systems on performance. They may be interested in examining those factors that influence (or have the strongest effect) on performance. This would require a different design.

between incentive systems (X) and years of experience (Z), and its effect (XZ) on performance. Exhibit 6-5 presents the research design.

The study can use two statistical methods of analysis. Campbell and Stanley (1963) recommend analysis of covariance for the proposed design. The matching variables are used as covariates. Regression analysis is also possible. However, Campbell and Stanley (p. 71) warn that even with a covariance analysis, "a significant treatment effect is interpretable only when *all* of the jointly contributing matching variables have been included".

Exhibit 6-5
Research Design

		GROUP 1	GROUP 2	GROUP 3
INDEPENDENT VARIABLES	TREATMENT VARIABLE (incentives system)	X_1	X_2	X_3
	TREATMENT VARIABLE (years of experience, prior to recruitment)	Z	Z	Z
	INTERACTION (incentive-recruitment)	X_1Z	X_2Z	X_3Z
	MATCHED VARIABLES (No. of families-per-promoter; No. of promoters-per-supervisor; Geographical dispersion of catchment area (households per km^2); No. of clients per service type; No. of assigned tasks;)	Y_{i-j}	Y_{i-j}	Y_{i-j}
DEPENDENT VARIABLES	- Real coverage (number of visits/different HH/period) - Continuity of coverage (number of follow-up visits/HH/period) - Knowledge levels - Quality of service provision (to-be-determined)	0	0	0

Another issue is obtaining sufficient sample size for each block. Data for the matching variables can be obtained from the PVOs and PSI. The researcher will select promoters (and their communities) that can be matched according to the confounding variables (Y_{i-j}) that could affect performance. Then these communities will be grouped according to PVO/incentive system used. For each group of promoters/communities, a sample will be selected randomly. Within each community a sample of households will be selected.

6.2.2.3 Level of Effort

1 Senior Economist:	20 days	@ $310/day
1 Level 2 Analyst:	30 days	@ $183/day
1 Research Assistant	20 days	@ $120/day
1 trip (senior economist)		@ $7,363
2 trips Level 2 Analyst		@ $14,726

Local hires:		
2 Researchers	20 days (each)	@ $50-100/day
10 Enumerators	20 days (each)	@ $16/day
1 Supervisor	30 days	@ $50/day
1 coder	20 days	@ $16/day

per diem ($20/day)—local hires

Estimated Cost: $50,000

6.2.3 A Comparative Study of Public and Private Provider Incentive Systems in Hospitals in Egypt

The government of Egypt is planning to grant greater managerial and financial autonomy to some of its hospitals. The objectives for doing so are to increase efficiency in their operations and reduce the burden the hospitals represent on the public budget. As part of the preparations of this change, Egypt is planning case studies to compare the differences in how government and private hospitals are managed. One aspect of these studies is to determine what incentive systems are provided in the private sector and their effects on the performance of hospital workers. The results of these studies will be used to guide the changes that need to be introduced in public hospitals receiving greater autonomy.

The detailed methodology of these case studies will be defined during an upcoming visit by HFS's Technical Director to Cairo.

BIBLIOGRAPHY

BAIRD, Charles W. (1971), "On Profits and Hospitals," *Journal of Economic Issues*, March, 57-66.

BARNUM, Howard, Joseph KUTZIN (1990), "Public Hospitals in Developing Countries: Resource Use, Costs and Financing," World Bank, Population, Health and Nutrition Division, Population and Human Resources Department.

BERENSON, Robert A. (1986), "Commentary: Capitation and Conflict of Interest," *Health Affairs*, (Spring).

BITRAN, Ricardo A. (1992) "Technical and Economic Efficiency in the Production of Health Services," Health Financing and Sustainability (HFS) Project, Bethesda, MD.

BITRAN, R., Munkatu MPESE, Taryn VIAN, and others (1987) "Zaire Health Zone Financing Study," Resources for Child Health (REACH) Project, Arlington, VA.

BOVBJERG, Randall, Philip HELD, and Mark PAUL, "Privatization and Bidding in the Health-Care Sector," *Journal of Policy Analysis and Management*, vol. 6, no. 4, 648-666.

CAMPBELL, Donald T. and Julian C. STANLEY, (1963), *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin.

CLARKSON, Kenneth (1972), "Some Implications of Property Rights in Hospital Management," *Journal of Law and Economics*, (Oct., 363-84).

CLELAND, Catherine (1984), "Possibilities for HMO-Type Organizations in Less Developed Countries," Dept. of Health and Human Services, Health Resources and Services Administration, Office of HMOs.

CLEVERLY, William, and Robert MULLEN (1982), "Management Incentive Systems and Economic Performance in Health Care Organizations," *Health Care Management Review*, (Winter).

COBURN, Catherine d. Crone (1984), "Management Incentive Compensation in Hospitals: A Profile of Selected Cases," unpublished master's thesis, Sloan School of Management, MIT.

CONTANDRIOPOULOS, André-Pierre, François CHAMPAGNE, and Enis BARIS (1993), "Physician Compensation and Health Care System Objectives: An Appraisal of International Experiences," Groupe de recherche interdisciplinaire en santé, Faculté de médecine, University of Montreal, mimeo.

EGDAHL, Richard, and Cynthia TAFT (1986), "Financial Incentives to Physicians," *New England Journal of Medicine*, vol. 315, no. 1.

ELLIS, Randall and Mukesh CHAWLA (1992), "Public and Private Interactions in the Health Sector," Draft, Health Financing and Sustainability Project, Bethesda, MD: HFS.

FREUND, Deborah A. (1987), "Competitive Health Plans and Alternative Payment Arrangement for Physicians in the United States: Public Sector," *Health Policy*, 7, 163-173.

_____. Richard SHACHTMAN, Marshall RUFFIN, Dana QUADE (1985), "Analysis of Length-of-Stay Differences Between Investor-Owned and Voluntary Hospitals," *Inquiry* (Spring).

GOULET, Denis (1989), *Incentives for Development: The Key to Equity*. New York: New Horizons Press.

GROUP HEALTH ASSOCIATION OF AMERICA, INC. (1985), "Managed Prepaid Health Care in Latin America and the Caribbean: A Critical Assessment," USAID, Bureau for Latin America and the Caribbean.

HARRIS, Jeffrey (1977), "The Internal Organization of Hospitals: Some Economic Implications," *The Bell Journal of Economics*, Autumn, vol. 8, no. 2.

HEMENWAY, David, et. al. (1990), "Physicians' Responses to Financial Incentives: Evidence from a For-Profit Ambulatory Care Center," *New England Journal of Medicine*, vol. 322, no. 15.

HILLMAN, Alan (1987), "Financial Incentives for Physicians in HMOs: Is There a Conflict of Interest?", *New England Journal of Medicine*, vol. 317, no. 27.

_____ and Mark PAULY (1989), "How Do Financial Incentives Affect Physicians' Clinical Decisions and the Financial Performance of Health Maintenance Organizations?", *New England Journal of Medicine*, vol. 321, no. 2.

KIRKMAN-LIFF, Bradford, and Wynand P.M.M. VAN DE VEN (1989), "Improving Efficiency in the Dutch Health Care System: Current Innovations and Future Options," *Health Policy*, 13, 35-53.

JONES, Leroy P. (1991), "Performance Evaluation for Public Enterprises," World Bank Discussion Paper 122, Washington, D.C.: World Bank.

KNIPPENBERG, Rudolph, et al., (1990), "The Bamako Initiative: Some Experiences...," Children in the Tropics, Review of the International Children's Center (CIE), no. 184/185. (micro-planner).

KOURI, Yamil, Donald SHEPARD, Freddie BORRAS, Jeannette SOTOMAYOR, George GELLERT (1991), "Improving the Cost-Effectiveness of AIDS Health Care in San Juan, Puerto Rico," *The Lancet*, 337: 1397-99.

LEE, Maw Lin (1971), "A Conspicuous Production Theory of Hospital Behavior," *Southern Economic Journal*, (July).

LEWIS, Maureen A., Margaret B. SULVETTA and Gerard M. LAFORGIA. (1990). "Estimating Public Hospital Costs by Measuring Resource: A Dominican Case." The Urban Institute, Washington, D.C., no. 3714-06. (July).

LEWIS, Maureen (1988), "Financing Health Care in Jamaica," Washington, D.C.; The Urban Institute 3714-04.

MANAGEMENT SCIENCES FOR HEALTH (1989), "Introduction: PROSALUD's Operations Research Studies," USAID/Bolivia.

_____ (1989a), "Employee Incentive System," PROSALUD, USAID/Bolivia.

_____ (1989b), "Financing Primary Health Care: Lessons from Bolivia," PROSALUD, USAID/Bolivia.

MAYNARD, Alan (1987), "Incentives for Cost-Effective Physician Behavior," *Health Policy*, 7, 189-204.

MUSGROVE, Philip (1986), "Measurement of Equity in Health," *World Health Statistics Quarterly*, 39, p.235.

OGAR, Jonathan (1974), "The Nonprofit Firm: A Test of the Theory for the Hospital Industry," *Journal of Economics and Business*, (Winter), 115-23.

PATTISON, Robert, Hallie KATZ (1983), "Investor-Owned and Not-For-Profit Hospitals: A Comparison Based on California Data," *New England Journal of Medicine*, vol. 309, no. 6.

PAULY, Mark V. (1980), *Doctors and Their Workshops: Economic Models of Physician Behavior*. Chicago: University of Chicago Press.

_____ (1987), "Nonprofit Firms in Medical Markets," American Economics Association Papers and Proceedings, vol. 77, no. 2.

_____ (1970), "Efficiency, Incentives and Reimbursement for Health Care," *Inquiry*, vol. VII, no. 1.

_____ and Michael REDISCH (1973), "The Not-For-Profit Hospital as a Physicians' Cooperative," *American Economic Review*, vol. 63, no. 1.

REGISTER, Charles A. and Edward R. BRUNNING (1987), "Profit Incentives and Technical Efficiency in the Production of Hospital Care," *Southern Economic Journal*, volume 53, pp. 899-914.

RELMAN, Arnold (1988), "Salaried Physicians and Economic Incentives," *New England Journal of Medicine*, Sept. 22.

RENN, Steven C., Carl SCHRAMM, J. Michael WATT, Robt. DERZON, (1985), "The Effects of Ownership and System Affiliation on the Economic Performance of Hospitals," *Inquiry* 22: 219-236 (Fall).

SHEPARD, Donald, Eckhard KLEINAU, J.M.V. RWABUKWISI (1991), "User Fees in Rwanda: An Empirical Plan," Cambridge, Harvard Institute for International Development.

_____ Taryn VIAN, Eckhard KLEINAU (1990), "Health Insurance in Zaire," World Bank, Africa Technical Department Working Paper 489.

SOLARI, A. (1984) "HMO and/or Prepaid Health Activity in South American Countries.", mimeo, Montevideo, Uruguay.

STEVENS, Carl M. (1991), "Egypt Cost Recovery Programs in Health Project Component One (Cost Recovery Hospitals): Project Design and Implementation," Health Financing & Sustainability (HFS) Project, Bethesda, MD.

_____ (1990), "Egypt Cost Recovery Programs in Health Project Component One (Cost Recovery Hospitals): Project Design and Implementation," (draft), USAID/Cairo.

_____ (1984), "Alternatives for Financing Health Services in Kenya."

_____ (1986), "Increasing the Efficiency of Health Services in Indonesia: A Key Strategy for Child Survival," USAID/Indonesia.

_____ (1989), "Indonesia: Health Sector Financing Project," USAID/Indonesia.

STIGLITZ, Joseph, (1986) *Economics of the Public Sector*. New York: W.W. Norton & Co.

TILNEY, J., Jr., R. BITRAN, D. DEAL, and B. BA, (1992) "The Gambia Review of Ministry of Health Cost Recovery Project," Abt Associates Inc., Cambridge, Massachusetts.

WELCH, W. Pete (1990), "Giving Physicians Incentives to Contain Costs Under Medicaid," *Health Care Financing Review* (Winter), vol. 12, no. 2.

WLODARCZYK, W. Cezary (1987), "In Search of Economic Rationality: The Experience of the Polish National Health Service," *Health Policy*, 7, 149-162.

WOUTERS, Annemarie (1990), "The Cost and Efficiency of Public and Private Health Care Facilities in Ogun State, Nigeria," Interdepartmental Program in Public Health Economics, Working Paper No. 4, The Johns Hopkins University School of Hygiene and Public Health, Baltimore, MD.

WYSZEWIANSKI, Leon, J. William Thomas, and Bruce A. Friedman (1987), "Case-Based Payment and the Control of Quality and Efficiency in Hospitals," *Inquiry*, 24: 17-25, (Spring).

APPENDIX

LIST OF PERSONS CONTACTED

John Akin (*University of North Carolina*)

Howard Barnum (*The World Bank*)

William Bicknell (*Boston University*)

Tina Clayland (*USAID*)

Catherine Crone Coburn (*Management Sciences for Health*)

Allan Fairbank (*Congressional Budget Office*)

Charles Griffin (*HFS Project*)

Karen Lashman (*The World Bank*)

Maureen Lewis (*The World Bank*)

Marty Makinen (*Abt Associates*)

William Marder (*Abt Associates*)

Lynn Miller Franco (*University Research Corporation*)

Philip Musgrove (*The World Bank*)

Mead Over (*The World Bank*)

Victor Pakqueo (*The World Bank*)

David Parker (*UNICEF*)

Gerry Rosenthal (*John Snow*)

Judy Seltzer (*Management Sciences for Health*)

Donald Shepard (*Harvard Institute for International Development*)

Carl Stevens (*Reed College*)

Wayne Stinson (*University Research Corporation*)

Taryn Vian (*Management Sciences for Health*)

Annemarie Wouters (*Johns Hopkins School of Hygiene and Public Health*)